Prioritization of Quality Factors in Agent Based Web Services

THESIS

Submitted

In fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

By

Amandeep Kaur

University Regd. No PHDENG10031

Supervised by

Dr. Rachna Soni Head Deptt. of Computer Science and Applications DAV College for Girls, Yamuna Nagar, Haryana Dr. Sudhir Mahajan Professor Deptt. of Computer Science and Engg. Chitkara University, Himachal Pradesh

June, 2016



Department of Computer Science and Engineering

CHITKARA UNIVERSITY, HIMACHAL PRADESH, INDIA

Prioritization of Quality Factors in Agent Based Web Services

THESIS

Submitted

In fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

By

Amandeep Kaur

University Regd. No. PHDENG10031

Supervised by

Dr. Rachna Soni Head Deptt. of Computer Science and Applications DAV College for Girls, Yamuna Nagar, Haryana Dr. Sudhir Mahajan Professor Deptt. of Computer Science and Engg. Chitkara University, Himachal Pradesh

June, 2016



Department of Computer Science and Engineering

CHITKARA UNIVERSITY, HIMUDA EDUCATIONAL HUB, SOLAN, HIMACHAL PRADESH-174103

CHITKARA UNVERSITY, HIMACHAL PRADESH

DECLARATION BY THE STUDENT

I hereby certify that the work which is being presented in this thesis entitled "Prioritization of Quality Factors in Agent Based Web Services" is for fulfilment of the requirement for the award of Degree of Doctor of Philosophy submitted in the Department of Computer Science and Engineering, Chitkara University, Barotiwala, Solan, Himachal Pradesh is an authentic record of my own work carried out under the supervision of Dr. Rachna Soni and Dr. Sudhir Mahajan.

The work has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgements have been made wherever the findings of others have been cited.

(Amandeep Kaur)

CHITKARA UNVERSITY, HIMACHAL PRADESH

CERTIFICATE BY THE SUPERVISOR(S)

This is to certify that the thesis entitled "**Prioritization of Quality Factors in Agent Based Web Services**" submitted by **Amandeep Kaur, Regd. No. PHDENG10031** to the Chitkara University, Barotiwala, Solan,Himachal Pradesh in fulfilment for the award of the degree of **Doctor of Philosophy** is a *bona fide* record of research work carried out by her under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institution or University for the award of any degree or diploma.

Dr. Rachna Soni Head Deptt. of Computer Science and Applications DAV College for Girls, Yamuna Nagar, Haryana

Dr. Sudhir Mahajan Professor Deptt. of Computer Science and Engg. Chitkara University, Himachal Pradesh

ACKNOWLEDGMENT

I take this opportunity to thanks **Brg. (Dr.) RS Grewal**, **Vice Chancellor**, **Dr. Varinder Kanwar**, **Registrar**, **Dr. Rajnish Sharma**, **Academics Dean**, **Dr. Rachna Soni**, **Head of Deptt. of Computer Science and Applications**, **Dr. Sudhir Mahajan**, **Professor**, **Deptt. Of Computer Science and Engg.**, **Research Committee** members and **Research Scholars** who helped in preparing these guidelines.

I extend my sincere thanks to one and all of Chitkara faculty for the completion of this document on the thesis format guidelines.

(Amandeep Kaur)

LIST OF PUBLICATIONS

Title of paper	Publication	Status	Year
A Technical Review on Agent Based Communication through Web Ontology	International Journal of Advanced Research in Computer Science and Software Engineering ISSN No.2277 128X	Published	2013
Semantic Languages in Web Services-A Review	Two days national workshop on computational Sanskrit-Issues and Challenges ,proceedings with ISBN No. 978-93-82529-03-0	Published	2013
Use Case Model of QoS with Extensible Registry Protocol	IEEE Explore,DOI Number 10.1109/CSNT.2015.2777	Published	2015
Role Of Classification Algorithms For Information Retrieval To Support Web Services Model	Journal Of Theoretical And Applied Information Technology (E-ISSN 1817- 3195 / ISSN 1992-8645) Scopus Indexed	Accepted	2016
Flaws in Determining QoS of Web Services	Cosmic Journal ,International Journal of Computer Science And Technology (IJCST)	Accepted	2016
Performance Analysis of different classification algorithms in Information Retrieval through Web Services	IJCST,Scopus Indexed	Accepted	2016
Innovative approach of data extraction in RDF using Symfony Framework	IJCST,Scopus Indexed	Waiting	2016

- HTTP Hypertext Transfer Protocol
- LDA Linear Discriminant Analysis
- NN Neural Network
- QOS Quality of Service
- **REST-Representational State Transfer Protocol**
- **RPC Remote Procedure Calls**
- SWS Semantic Web Services
- SVM Support Vector Machines
- SOAP Simple Object Access Protocol
- UDDI Universal Description, Discovery and Integration
- WSDL Web Service Description Language
- XML Extensible Mark-up Language

LIST OF TABLES

Table No.	Title	Page No).
Table 3.1	Tools used	46	5
Table 6.2	A List of 10 Jewellery Web Services before Rank in UDDI Registr	y 99)
Table 6.3	A List of 10 Jewellery Web Services after Rank in UDDI Registry	10)()
Table 6.4	Experimental results for algorithms Greedy, CloudRank, PSORank Bayes	and Nai	ive 03
Table 6.5	Result Analysis of accuracy evaluation	10)4
Table 6.6	Recall and Precision rate values w.r.t. SVM	10)7
Table 6.7	Recall and Precision rate values w.r.t. Fuzzy Logic	10)8
Table 6.8	Recall and Precision rate values w.r.t. LDA	11	.0
Table 6.9	Recall and Precision rate values w.r.t. NN	10)9
Table 6.10	Comparison between Traditional Methods	11	.0
Table 6.11	SVM Comparison	11	.2
Table 6:12	Fuzzy Comparison	11	.4
Table 6:13	LDA Comparison	11	.6
Table 6.14	Neural Network Comparison of Throughput and Response Time	11	.8
Table 6.15	Comparison of Proposed Model With Existing Model	12	20

LIST OF FIGURES

Figure No.	Title	Page No.
Figure 1.1	Web Service Model	18
Figure 3.1	Web Service Model	42
Figure 3.2	Proposed Model	43
Figure 3.3	Workflow of Proposed Model	45
Figure 3.4	Symfony Framework 3.0.0	47
Figure 3.5	Symfony Framework Performance Metrics	50
Figure 3.7	Command History	50
Figure 3.8	Workspace	50
Figure 3.9	Current Folder	50
Figure 3.10	Start GUI	51
Figure 3.11	NEW GUI	51
Figure 3.12	Other Options for GUI	52
Figure 3.13	To Draw a GUI	52
Figure 3.14	Editor Window	53
Figure 3.15	Fuzzy logic in MATLAB	54
Figure 3.16	Support Vector Machine	54
Figure 3.17	Neural Network	57
Figure 3.18	Neural Network Model	58
Figure 3.19	BPNN Process Flowchart	60
Figure 3.20	Naïve Bayes Classifier	61
Figure 3.21	LDA Flowchart	62
Figure 4.1	Simple Form of an E-Commerce Application	65
Figure 4.2	Interaction of Web Server with Multiple Data Bases	66
Figure 4.3	Role of Broker with Multiple Web Servers and Data Bases	66
Figure 4.4	Enhanced Use Case Model of Rating similar web services	67
Figure 4.5	Naive Bayes Implementation Flowchart	69
Figure 4.6	Sequence Diagram of Suggest Action	71
Figure 4.7	Sequence Diagram for Conversion from HTML Text to Plain Text	72

Figure 4.8	Flowchart shown the Role OF SEO in RDF Search Engine	73
Figure 5.1	MATLAB GUI	78
Figure 5.2	Proposed Work Flow in MATLAB	79
Figure 5.3	Snap Shot for uploading web page	80
Figure 5.4	Snapshot of data uploaded successfully under particular category	80
Figure 5.5	Snapshot showing removed stop words	81
Figure 5.6	Snap Shot showing weight matrix	82
Figure 5.7	Trained Data by SVM	83
Figure 5.8	Trained data using LDA	84
Figure 5.9	Trained Data Set using Fuzzy Logic	85
Figure 5.10	Trained Data using Neural Network	86
Figure 5.11	Search Accuracy and Error Percentage score by SVM	88
Figure 5.12	Throughput, Content availability and Response Time by SVM	88
Figure 5.13	Ranked result on the basis of score of various QoS parameters	90
Figure 5.14	Throughput, Content availability and Response Time by LDA	90
Figure 5.15	Search Accuracy and Error Percentage score by LDA	91
Figure 5.16	Ranked result on the basis of score of various QoS parameters	91
Figure 5.17	Throughput, Content availability and Response Time by Fuzzy Logic	93
Figure 5.18	Search Accuracy and Error Percentage score by Fuzzy Logic	93
Figure 5.19	Ranked result on the basis of score of various QoS parameters	94
Figure 5.20	Throughput, Content availability and Response Time by Neural Network	95
Figure 5.21	Search Accuracy and Error Percentage score by Neural Network	96
Figure 5.22	Ranked result on the basis of score of various QoS parameters	96
Figure 6.1	Comparison of Proposed Model With Existing Model	101
Figure 6.2	Snapshot of similar web services after ranking in Symfony Framework	101
Figure 6.3	Snapshot shown Overall Response Time and Throughput of 10 Jewellery Services	Web 102
Figure 6.4	Snapshot shown Individual Quality Factor Scores of 10 Jewellery Services	Web 102
Figure 6.5	Comparison of Response Time value between Existing and Proposed M	ethod 103
Figure 6.6	Comparison of Throughput value between existing and proposed method	104

Figure 6.7	Comparison of precision value between existing and proposed approaches	105
Figure 6.8	Comparison of Recall value between existing and proposed approaches	105
Figure 6.9	Recall and Precision Rate Graph using SVM	107
Figure 6.10	Recall and Precision Rate Graph using Fuzzy	110
Figure 6.11	Recall and Precision Rate Graph using LDA	110
Figure 6.12	Recall and Precision Rate Graph using NN	110
Figure 6.13	Throughput Comparison Graph w.r.t. SVM	113
Figure 6.14	Response Time Comparison Graph w.r.t. SVM	113
Figure 6.15	Throughput Comparison Graph w.r.t. Fuzzy	115
Figure 6.16	Response Time Comparison Graph w.r.t. Fuzzy	115
Figure 6.17	Throughput Comparison Graph w.r.t. LDA	117
Figure 6.18	Response Time Comparison Graph w.r.t. LDA	117
Figure 6.19	Throughput Comparison Graph w.r.t. NN	118
Figure 6.20	Response Time Comparison Graph w.r.t. NN	120

TABLE OF CONTENTS

Declaration	i
Certificate	iii
Acknowledgement	V
List Of Publications	vi
Abbreviations	. vii
List Of Tables	viii
List Of Figures	xi
Table Of Contents	. xii
Abstract	xvi
Chapter 1: Introduction	. 17
1.1.1 What Are Web Services	18
1.1.2 Properties of Web Services	.18
1.1.3 Web Service Model	.18
1.1.4 Web Service Description Languages	.19
1.1.5 The UDDI Registry	.20
1.1.6 Web Service Discovery	.20
1.2 Role of Semantic Web in Quality of Service	22
1.2.1 QoS Ontology	.23
1.2.2 Purpose Of Semantic Web	.26
1.2.3 Semantic Web Technologies	.27
1.3 Agent Based Semantic Web Services	28
1.4 Importance of Research	28
Chapter 2: LITERATURE SURVEY	

2.1 Related Work	
2.2 Inferences drawn out of Literature Survey	
2.3 Issues in Literature Survey	40

Chapter 3: Proposed Work And Methodology
3.1 Problem Statement And Objectives
3.2 Proposed Work
3.3 Methodology
3.4 Tools Used In Methodology 46
3.4.1 Symfony Framework
3.4.1.1 Installation
3.4.1.2 Twitter Bootstrap
3.4.2 Matlab
3.5 Classification Algorithms
3.5.1 Fuzzy Model
3.5.2 Support Vector Machine (SVM)
3.5.2.1 Basic Operations Utilize The SVM
3.5.3 Neural Network
3.5.4 Naive Bayes Classifier
3.5.5 LDA61
Chapter 4: NAÏVE BAYSED APPROACH IN SYMFONY FRAMEWORK FOR
4.1 E COMMEDCE IN WED SEDVICES 63
4.1 E-COMIMERCE IN WEB SERVICES
4.2 PROPOSED MODEL FOR PRIORITIZATION OF AGENT BASED E-COMMERCE WEB SERVICES
4.2.1 Role Of Search Agent In Broker
4.2.1.1 Ranking Action (Using Naïve Bayes Approach)
4.2.1.2 Suggest Action
4.2.1.3 HTML Parser
4.2.2 Role Of SEO (Search Engine Optimization)

	4.2.3 Role Of Quality Manager In Qos Prioritization Model
	4.2.4 Role Of Service Provider In Qos Prioritization Model
4	4.3 Pros And Cons Of Naive Bayes
4	4.4 Protocol And Technology Advantages
Ch	apter 5: DATA CLASSIFICATION AND RANKING USING MATLAB
:	5.1 Select Category And Upload Data
	5.2 Remove Stopping Words
	5.2.1 Feature Generation
	5.2.2 Generator Used In Feature Generation
:	5.3 BUILD WEIGHT MATRIX
:	5.4 TRAINING DATA
	5.4.1 Training Data Using SVM(Support Vector Machine)
	5.4.2 Training Data Using LDA (Linear Discriminant Analysis)
	5.4.3 Training Data Using Fuzzy Logic
	5.4.4 Training Data Using Neural Network
	5.5 Data Classification And Ranking After Calculations Of Scores Of Quality Factors Using Algorithms SVM, Neural, LDA And Fuzzy
	5.5.1 Data Classification And Ranking Using SVM
	5.5.2 Data Classification And Ranking Using LDA
	5.5.3 Data Classification And Ranking Using Fuzzy Logic
	5.5.4 Data Classification and Ranking using Neural Network94
Ch	napter 6: RESULTS AND ANALYSIS
(6.1 Results Using Naïve Based Algorithm In Symfony Framework
	6.1.1 Performance Analysis Using Naïve Baysed Classifier
(6.2 Results using other Data Classification Algorithms in MATLAB 106
	6.2.1 Results and Analysis using Recall Rate and Precision Rate in MATLAB
	6.2.2 Results and Analysis using Throughput and Response Time in MATLAB

6.3 COMPARISION	
Chapter 7: SUMMARY AND OUTLOOK	
7.1 SUMMARY	
7.2 OUTLOOK	
REFERENCES	

ABSTRACT

As the utilization of web is expanding more step by step, the web client get easily lost in the web's rich hyper structure. The primary point of proprietor of the website is to give the important data in terms of satisfactory QoS (Quality of Service) factors. From the client point of view, Web Service based QoS Discovery is a multi-criteria decision mechanism that requires knowledge about the service and its QoS description. These clients are not experienced enough to acquire the best selection of web service and trust the QoS information published by the provider. In the recent years, several resources have improved the web framework making it possible for it to include and exhibit other media types, such as pictures, movies and sound, thus making the user interaction more interesting. To solve the problems encountered in literature review, many researchers attempt to make a rational and effective decision for QoS based service selection. The new discovery approach is suggested to be the solution for the current web service discovery problems. The existing tModel was used with XML based SOAP protocol in order to solve the problem of UDDI registry. In this research, tModel for storage format has been enhanced with the help of RDF (Resource Description Framework) which supports JSON format. Widespread access to internet has facilitated the corporate houses to enhance their customer base through e-commerce websites/ portals. A comparative performance analysis of prominent page rank algorithms was made on the basis of metrics like throughput, responsetime, recall rate and precision rate etc. The experiment revealed the fact that recall and precision rate are the best to predict the Quality of Service (OoS) supported by various E-Commerce web sites like Amazon, Jabong and Shopclues etc. Simulation Interface has been designed for classification algorithms. The program is developed for the Fuzzy Logic, Naïve Bayes, Neural Network, Linear Discriminant Analysis and Support Vector Machine using Symfony Framework and MATLAB .The primary reason is to examine the classification algorithms utilized for information retrieval in terms of satisfactory QoS factors. Detailed performance analysis further concluded that Neural Network could be the best algorithm to rate the service quality of E-Commerce websites.

WWW gives Information System to achieve data representation, data recovery, assignments and information administration. With the end goal of data recovery, WWW gives Web Services and Web applications. Web applications incorporate Internet specialized tools, informal organization and E-Commerce [1]. Web Services assuming a critical part in the E-Business, Computer Software and Communication Industries. Semantic Web Services (SWS) as coordinated answer for understanding the vision of the up and coming era of the web [2]. These administrations are self-depicted and give important and interpretable data of each web content so that individuals and PCs empower to work in participation. An agent is one who talks on the behalf of other entities too [3]. Agent Systems are independent programming programs handling space information and a capacity to act with some level of freedom to complete activities to accomplish determined objectives. Web Services are application that impart utilizing open conventions, for example, Hypertext Transfer Protocol (HTTP), Extensible Mark-up Language (XML) and Simple Object Access Protocol (SOAP). They are intended to bolster machine-to-machine communication over a system. Numerous Companies give similar web services to clients. For instance, Google permits programming designers to question billions of site pages straightforwardly from their own PC programs. An engineer can utilize his or her most loved programming language, for example, Java or Visual Studio.Net to create applications that get to the Google Web Services. Applications can be changed over into web applications through Web Services which can be found, distributed and utilized through the web and any application can utilize web administrations. As appeared in Web Service Model in Figure 1, Web Services depend on the functionalities of distribute, find and tie exercises and the segments of basic web service model are: Service Provider, Service Broker and Service Requester. Service Provider distributes administrations in searchable registry or catalogue. Service Requester is customer asking for administration from administration supplier through Service Broker which goes about as a centre product between Service Requester and Service Provider.



Figure 1.1 Web Service Model

Above Figure 1.1 demonstrated is the base model (Web Service Model) that has the administration supplier that chooses administration by utilizing the administration supplier. Administration supplier distributes the depiction of the administration capacities. Correspondence between elements of web administration model is done through correspondence conventions.

1.1.1 WHAT ARE WEB SERVICES?

Web Services are the languages for writing the web applications having XML, UDDI, and SOAP as their components. XML is used for tagging, UDDI is used to store information and SOAP is used for exchanging of the information.

1.1.2 PROPERTIES OF WEB SERVICES

FUNCTIONAL PROPERTIES-Functional Properties mean what the service can do. These describes behaviour of the web services. Any layman can observe the behaviour of web services but to know about the back ground or hidden details of web services is impossible. **NON FUNCTIONAL PROPERTIES**-Non Functional means how the service can do. These describes the ways through which performance of Functional Properties can be improved. In short, how the service can perform to achieve a particular task. For example: Throughput, Response Time, Content availability, Accuracy.

1.1.3 WEB SERVICE MODEL

In a Web administration display, an administration supplier offers web administrations which give capacities or business operations which can be conveyed over the Internet, with the expectation that they will be conjured by accomplices or clients; a Web administration requester depicts necessities keeping in mind the end goal to find administration suppliers. Distributed, authoritative, and finding. Web administrations are three noteworthy errands in the model. Disclosure is the procedure of discovering Web administrations supplier areas which fulfill particular prerequisites. Web administrations are futile on the off chance that they can't be found. In this way, discovery is the most vital errand in the Web administration model. The Web administration model in Fig. 1.1 demonstrates the association between an administration requester, administration suppliers, and an administration disclosure framework. The service providers offer Web services which provide functions or business operations.

Web administration requesters and Web administration suppliers will impart by means of SOAP convention which is a XML based convention for Web administration trade data.

1.1.4 WEB SERVICE DESCRIPTION LANGUAGES

WSDL is a standard to portray how to get to a Web administration and what operations (techniques) it performs. The WSDL is a XML-based dialect, which indicates a web administration by characterizing messages that give a theoretical meaning of the information being transmitted and operations that a web administration gives to transmit the messages. WSDL gives a capacity driven depiction of covering inputs, yields and special case taking care of. It is additionally used to find Web Service when a customer requests and supplier use WSDL to depict the web service information. A WSDL archive characterizes a Web administration utilizing four noteworthy components, to be specific, port, message, sorts, and authoritative. An elaboration of these components is as per the following:

- WSDL Ports: The <portType> is the most imperative component in WSDL. It characterizes a web administration with operations that can be performed and messages that are included. The <portType> component is like a capacity library (or a module, or a class) in a customary programming dialect.
- WSDL Messages: The <message> component characterizes a unique, wrote meaning of the information being conveyed. Every message can comprise of one or more intelligent parts like parameters of a capacity bring in a conventional programming dialect.
- WSDL Types: The <types> component characterizes a holder for information sort definitions that are important for the traded messages. For most extreme interoperability and stage lack of bias, WSDL utilizes XML Schema sentence structure to characterize information sorts.

4. WSDL Bindings: The

binding> component characterizes a solid convention and information group detail for a specific port sort.

1.1.5 The UDDI Registry

UDDI was proposed by web services. It is the database of all the web services. Travel offices utilize UDDI to check their stored database. Also, administration suppliers must distribute the data of their administrations before they are accessible to shoppers. The registry proprietor chooses who has the power to distribute and redesign administrations data. An organization is not ready to distribute or redesign the data of administrations gave by another organization. The registry proprietor chooses what data can be distributed in the registry. UDDI is a case of this methodology. Brought together registries are fitting in static or a controlled situation where data does not change regularly.

1.1.6 WEB SERVICE DISCOVERY

Web Service Discovery is a procedure of finding services that most appropriate to client's solicitation as indicated by requester's necessity. One of the primary difficulties in finding best web service is that administration registries don't give enough question components to customers to explain legitimate administration inquiries that can address their issues. For instance, administration registries permit customers to perform basic pursuit questions, for example, looking by administration name, tModel, or business name. Be that as it may, finding applicable Web administrations couldn't be accomplished utilizing basic keyword based look methods especially as Web administrations multiply. Besides, separating Web administrations from each other utilizing watchword coordinating methods is unreasonable since minimal literary data is regularly given in administration disclosure interfaces. Finding web administrations includes three interrelated stages: 1-coordinating, 2-appraisal, and 3determination. In stage 1, the administration portrayal contribution by the designer is coordinated to that of an arrangement of accessible assets. In stage 2, the after effect of coordinating (commonly an arrangement of positioned web administrations) is evaluated and sifted by a given arrangement of criteria. In stage 3, administrations are really chosen for consequent modifying and consolidating with others. A discovery administration, which could be performed by either a shopper specialist or a supplier operator, is expected to encourage the disclosure procedure.

Discovery Mechanisms: As the quantity of Web Services about particular thing expands, the accomplishment of organizations will rely on upon administration revelation. The objective is to discover suitable Web benefits that match an arrangement of client necessities. A first portrayal of revelation components for administration suppliers is the match-production

process. It is the procedure of finding a fitting administration supplier for an administration requester through a center specialist. It incorporates the accompanying general strides: a) Service suppliers promote their capacities to center operators, b) center specialists store this data, c) an administration requester asks a center specialist whether it knows of administration suppliers best coordinating asked for abilities and d) the center operator, keeping in mind the end goal to answer, tries to coordinate the solicitation against the put away ads and returns a subset of put away administration suppliers' notices. WS Discovery components incorporate a progression of registries, records, indexes, operator based and Peer to Peer-P2P arrangements. The three driving methodologies on how a disclosure administration ought to be planned are: a registry, a record, or a distributed (P2P) framework.

Registry: A registry is a legitimate, halfway controlled archive of administrations data. Administration suppliers must distribute the data of their administrations before they are accessible to shoppers. The registry proprietor chooses who has the power to distribute and redesign administrations data. An organization is not ready to distribute or redesign the data of administrations gave by another organization. The registry proprietor chooses what data can be distributed in the registry. UDDI is a case of this methodology. Brought together registries are fitting in static or a controlled situation where data does not change regularly.

List: A record is an accumulation of distributed data by the administration suppliers. It is not legitimate and the data is not halfway controlled. Anybody or any organization can make their own record, which gathers data of administrations uncovered on the web for the most part utilizing web bugs. The data in a record could be outdated however can be checked before use. Google is a case of the record approach.

Inventories: Web Service indexes are the ruling innovative premise for WS Discovery components. They are particular storehouses which actualize a determination system as met schema. Specifically, preceding the UDDI standard, associations did not have a typical way to deal with distribute data about their items and web administrations for their clients and accomplices?

UDDI set up the primary uniform strategy that included points of interest for joining of officially existing frameworks and procedures between business accomplices. UDDI permits the undertakings to find and impart data to respect to the web administrations and other electronic and no electronic administrations that are enlisted in a registry. The prerequisites incorporate watchwords, part of the administration's name and persistence, with a specific end

goal to choose the appropriate administration through the consequences of the registry. The accessible inquiry apparatuses are extremely basic and don't contemplate any cross-relationships between web administrations and the subjective attributes of every web administration, driving the client to rehash the hunt from the earliest starting point utilizing new catchphrases. There are three sorts of data upheld by the inventory. These sorts included enrollment of white, yellow and green pages. White pages incorporate essential contact data and identifiers, for example, association name, address, contact data, and other remarkable ids.

P2P-based Solutions: P2P can be characterized as immediate correspondence or joint effort between PCs, where none are essentially customer or server, yet all machines are equivalents - peers. P2P system is PC system which depends essentially on the registering force and transfer speed of the members in the system instead of moving it in a generally low number of servers. An immaculate distributed system does not have the thought of customers or servers, but rather just equivalent companion hubs that all the while capacity as both "customers" and "servers" to alternate hubs on the system. P2P figuring gives a de-brought together option that permits Web Services to find each other progressively. Every Web administration is a hub in a system of companions. At revelation time, a Web Service questions its neighbors looking for an appropriate Web service. On the off chance that any of its neighboring companions coordinates its necessities, it answers and the inquiry is finished. Something else, the inquiry is engendered through the system until an appropriate Web administration is found or certain end criteria are come to. P2P design is more dependable than registry approach since it does not require to bother about brought together registry, yet presents more execution expenses subsequent to the majority of time a hub goes about as a re-layer of data.

A P2P overlay system gives a foundation to routing and information area in a decentralized, self-organized environment in which every companion demonstrations not just as a hub giving directing and information area administration, additionally as a server giving administration access. P2P can be viewed as a complete circulated registering model. Hubs are relegated IDs drawn from a worldwide location space. Companions are additionally doled out a scope of keys from the worldwide location space that they are in charge of.

1.2 ROLE OF SEMANTIC WEB IN QUALITY OF SERVICE

Semantic Web Services (SWS) as an incorporated answer for understanding the vision of the up and coming era of the web [2]. These administrations are self-portrayed and give significant and interpretable data of each web content so that individuals and PCs empower to work in participation. The cutting edge guarantees to convey Semantic Web Services (SWS)

that are self-depicted and amendable to robotized disclosure, creation and conjuring. Semantic Web gives the foundation to the semantic interpretability of web administrations. Semantic Web administrations give formal depiction of each idea in web administrations furthermore gives the relationship between ideas; so web administrations can be used by applications or different administrations without human help. In this manner, Semantic Web Services can possibly change the way information representation.

1.2.1 QOS ONTOLOGY

QoS metaphysics gives administration operators a chance to coordinate publicized quality levels for its shoppers with particular QoS inclinations. Suppliers express arrangements and customers express inclinations utilizing the QoS philosophy, which additionally empowers the buyers to design administration intermediary operators with the goal that they have the important practices to screen and record shopper and administration cooperation. It recognizes three ontologies for QoS: upper, center, and lower. The upper philosophy catches the most nonspecific quality ideas and characterizes the fundamental ideas connected with a quality, for example, quality estimation and connections:

- Quality speaks to a quantifiable non-functional part of a web service inside a given area.
 Quality credits identify with each other. The center cosmology contrasts from upper and lower ontologies in that it catches the quality ideas that are relevant to different areas, (for example, execution qualities).
- QAttribute catches a given quality's sort for instance, a quality whose worth is a floating-point number and increments in the quality reflect changes in the quality.
- QMeasurement measures a Quality unbiasedly or subjectively. Specialists make target estimations naturally, though subjective estimations include people. An estimation has a legitimacy period and can be affirmed.
- QRelationship portrays how qualities are associated. Administration reaction time, for instance, could be adversely related to throughput.
- AggregateQuality is a quality created from different qualities. The QoS center philosophy joins a few quality perspectives experienced in disseminated frameworks.
- Availability is the likelihood that an administration can react to buyer demands. It has two subclasses:
 - MTTR (mean time to repair, which means the normal time for reestablishing a fizzled administration) and

• UpTime (the span for which the administration has been operational consistently without disappointment).

Dependability

Dependability prerequisite measures the nature of a Web administration regarding how it performs, given a particular measure of time and the present system conditions, while keeping up the administration quality. It is additionally controlled by the quantity of disappointments every day and the medium of conveyance. Truth be told, Web Service Reliability (or WSReliability) is a most recent particular for open, dependable Web administration informing. WS-Reliability guarantees ensured conveyance of messages, end and/or identification of copy messages and the right requesting of messages. The WS-Reliability can be installed into SOAP as an extra augmentation as opposed to a vehicle level convention. This particular gives dependability notwithstanding interoperability, along these lines permitting correspondence in a stage and seller freeway. The WSReliability detail characterizes an arrangement of SOAP headers and guidelines in SOAP envelopes that deal with the message affirmations, message requesting and so on. WSReliability settles on offbeat informing a sound decision in light of the additional components that it gives like message affirmations and message requesting which thus permit the conveying gatherings to be more autonomous of each other. Dependability decides the rate of the times an occasion is finished with achievement. This numeral will give a chance to the administration purchasers to expect the likelihood of a disappointment that can happen amid an exchange. These disappointments depends on the quantity of dropped conveyances, copy conveyances, broken message conveyances, and out-of-request conveyances. An occasion may either succeed or fall flat, and there is no center ground in that issue. In this manner, absolute number of occasions will be the quantity of disappointments added to the quantity of fruitful occasions.

Adaptability

Adaptability prerequisite characterizes how expandable a Web Service Model can be. Any proposed web service model are being acquainted with new interfaces and strategies regular and this makes staying up with the latest a critical need. On the off chance that the circumstance requests for all the more processing capacities and overhauling more demands, the framework ought to be fit for supporting extra frameworks and more up to date advancements. The Web Service Model ought to have the capacity to handle substantial burden while ensuring that the execution as far as reaction time experienced by their customers is not offensive. The Performance Non-Scalability Likelihood (PNL) metric is a generally more current procedure to foresee whether the framework will be ready to withstand the higher heaps of movement without influencing the execution levels.

Exactness

It is the measure of the accuracy of web services. It is the recall rate and precision rate of the web administration.

Throughput

It is the amount of number of web services returned in a particular unit of time. It can be represented in percentage and in number of web services. According to Wikipedia, Throughput is a measure of how many units of information a system can process in a given amount of time.

Response Time

As per Wikipedia, Response Time is the total amount of time it takes to respond to a request for service

Respectability

Respectability prerequisite guarantees that any changes to proposed Web Service Model are performed in an approved way. Information honesty guarantees that the information is not ruined amid the exchange, and in the event that it debased, it guarantees that there are sufficient instruments in the configuration that can recognize such alterations. Information honesty is an imperative component to consider, in light of the fact that overlooking it might harm substantial programming modules and make mistakes that are difficult to follow back. The information messages that are gotten are checked to check whether they have not been changed in travel. This should be possible with procedures like checksum estimation or computerized marks. There are various instruments in the business sector like SIFT that can gather and screen the information being sent and got between the imparting parties. Information uprightness is a Boolean quality, implying that information either has trustworthiness the information holds. Information uprightness can hence be ascertained as the proportion of effective exchanges to the aggregate number of exchanges.

Accessibility

It guarantees the less measure of disappointments by clients. Gives the less chance for saying of down time.

Availability

Availability is regularly a measure of the achievement rate of an administration instantiation at a given time. A Web administration won't be open despite the fact that it is still accessible, on the grounds that a framework might be up and running yet won't not have the capacity to prepare a solicitation perhaps as a result of the heap it is encountering. Availability thusly relies on upon how versatile the Web administration framework outline is, on account of an exceedingly adaptable framework continually serves the solicitation independent of the volume of the Web administration demands. Openness is the proportion of the quantity of fruitful reactions got from the server to the quantity of solicitations messages sent by the customers. It can be portrayed as the level of a framework at which it is equipped for reacting to the client summons of the administration. Regardless of kind of affirmations got, (either negative/positive or right/mistaken), availability can be figured as a proportion of number of fruitful affirmations got to the aggregate number of solicitations sent.

Interoperability

Interoperability implies that a Web Service can be utilized by any framework, regardless of working framework or framework design and that exact and indistinguishable result is rendered in any environment. Interoperability is a measure of how adaptable the Web administration has been made so that the customers don't need to stress over authoritative to a Web administration that can't be keep running in their surroundings. The formative environment here incorporates working framework, programming dialect or equipment sort. The between operability can be figured as the proportion of the aggregate number of situations the Web administration rushes to the aggregate number of conceivable situations that can be utilized.

This interoperability esteem measures the effective execution of the Web administration in various situations, (for example, diverse working frameworks, programming dialects, and equipment sorts). The quality more like one demonstrates higher interoperability which is alluring.

1.2.2 PURPOSE OF SEMANTIC WEB

The motivation behind Semantic Web is driving the development of the web by empowering clients to discover ,share and consolidate data all the more effortlessly .Humans are fit for utilizing the web to complete undertakings, for example, finding the ABC word for "organizer", saving a library book, and hunting down the most reduced value DVD. Be that as it may, machines cannot perform these errands without human heading, since website pages are intended to be perused by individuals, not machines. Semantic Web gives the fundamental strategies to thinking about each idea. It empowers semantic interoperability of web administrations through the distinguishing proof of semantically comparative ideas. It portrays depiction about each web benefit with the goal that product can consequently decide its motivation. Semantic portrayals of web administrations are essential with a specific end goal to empower their programmed revelation, organization and execution crosswise over heterogeneous clients and spaces. Existing Technologies for web benefits just give depictions

at the syntactic level, making it troublesome for requesters and suppliers to translate or speak to nontrivial explanations, for example, the significance of inputs and yields. This confinement might be casual by giving a rich arrangement of semantic comments that enlarge the administration portrayal. Organizations can speak to the conceivable ways they can connect as semantic and syntactic requirements.

1.2.3 SEMANTIC WEB TECHNOLOGIES

The Semantic Web is a dream of a Web of important substance and administrations, which can be translated by PC programs. It can likewise be seen as an unlimited wellspring of data, which can be displayed with the reason for sharing and reusing information. Semantic Web clients will have the capacity to accomplish more exact pursuits of the data and the administrations they require from the instruments gave. The Semantic Web gives the important base to distribute and determining ontological depictions of terms and ideas. Key to the interoperation of Web Services is a reception of an arrangement of empowering standard conventions. A few XML based principles have been proposed to bolster correspondence over WWW. XML outline (XML-S) [4] gives the hidden structure for both characterizing the Web Services Standards, a variables/objects/information sorts and so forth that are traded between administrations. Cleanser is W3C's suggested XML-information transport convention, utilized for information trade over online interchanges protocols(http).SOAP messages can convey a XML payload characterized utilizing XML-S, in this manner guaranteeing a steady elucidation of information things between various administrations.

WSDL is the W3C prescribe dialect for portraying the administration interface. RDF is XML based standard from W3C for depicting assets on the Web.RDF acquaints a little semantics with XML information by permitting the representation of items and their relations through properties. RDF-Schema is a straightforward sort framework, which gives data to the elucidation of the announcements given in RDF information. Web Ontology Language will give more prominent machine interpretability of Web substance than RDF and RDF Schema by giving a much wealthier arrangement of develops for determining classes and relations.

Ontologies have been produced inside the Knowledge Modelling research group [6] with a specific end goal to encourage information sharing and reuse. They give more prominent expressiveness when demonstrating area information and can be utilized to impart this learning amongst individuals and heterogeneous and conveyed application frameworks.

1.3 AGENT BASED SEMANTIC WEB SERVICES

There are numerous current meanings of agents in the writing. However an expanding number of specialists locate the accompanying characterisation helpful [5]: An operator is an epitomized PC framework that is arranged in some environment and that is equipped for adaptable, self-governing activity in that environment keeping in mind the end goal to meet its configuration targets. Programming specialist as an agent is a principle segment in building semantic web .It goes about as a middleware to associate with web services in the interest of clients. The semantic web characterizes and connects web services in a way that it can be utilized by machines with the end goal of computerization, combination and reuse of information over different applications in the form of different agents. The programming specialists can work successfully as more machine-discernible web content and mechanized web administrations are accessible. Agent Technology for web applications is confronting ceaseless changes. Numerous web advances have been produced and joined with web specialists as agents. A Semantic Web would permit fake agents to take the necessary steps of hunting down and arranging administrations required by people or associations. Current work on the Semantic Web and Web Services are personally laced with work on operator based frameworks. As a result of the self-governance nature of operators, specialist arranged procedures show potential in the Semantic Web, where operators are required both to give benefits and to make best utilization of the assets accessible.

1.4 IMPORTANCE OF RESEARCH

Web Services playing an important role in E-Business, Computer Software and Communication Industries. Problem arises when different web services return similar results. User feel similar functional properties (what the service can do), but in reality non-functional (how the system behave) properties i.e.; Throughput, Response Time, Accuracy rate vary. Many author attempted to solve the same problem but still technical issues, security issues, protocol issues, registry issues, model adaptable issues need to be concentrated. This Research gap derives Research Problem and solution of research will enables web clients to make comparison between similar web services in terms of Throughput, Response Time, and Accuracy rate. The proposed model will acts as a helping tool for the client to select best web service among similar web services in terms of their quality factors in eco-friendly environment. Service Consumer who is keenly interested in web services quality parameters may select a good quality and secure web service with less timing. Suggested helping tool will help consumers to distinguish white from black and comparisons can be made among similar services producing similar results. Hence, society interested in QoS of web services will find a good support while web discovery following proposed model.

In Chapter-2, Literature review has been discussed.

Research Problem, Research Objectives, Proposed Model and Methodology as well as various tools like Symfony Framework and MATLAB are mentioned in **Chapter-3**.

Naïve Baysed Implementation of proposed model has been shown in Chapter-4.

In **Chapter-5**, Implementation of proposed model is shown with the help of SVM (Support Vector Machine), LDA (Linear Discriminant Analysis), Fuzzy Logic and Neural Network.

In **Chapter-6**, Results and Analysis with Naïve Bayes has been presented and comparison has been made with existing approaches as well as Results and Analysis with classification algorithms has been shown and finally the results of SVM, LDA, Fuzzy Logic and Neural Network has been compared and found the best one.

In Chapter-7, Summary and Scope has been presented

Web services normally comprise of XML, SOAP and UDDI for achieving its features. XML for tagging the data, UDDI list down the web services which are available and SOAP used for shifting the data. By the extensive propagation of the Web Services, QoS (Quality of Service) has become an important factor in distinguishing the success of service providers. QoS has now a challenging task due to the dynamic and unpredictable character of the web. Mostly, the issues that are not solved can create critical transactional applications to endure by the improper levels of performance deprivation. It has been shown that the QoS-aware programs can have a higher business value as compare to other web services.

2.1 RELATED WORK

Daniel R.M. et al.[1996][6] proposed the method for text categorisation based on the usage of the data mining. In proposed work –means clustering and Euclidian distance has been proposed for text categorization.

Valentine et al.[1997][7] proposed the usage of hyperlinks for analysing pages of the packages. This research provides a traceable process for extracting hierarchal model for generating a large real life applications. They follows Semantic Modelling and Development tools like XML.But these XML based tools are insecure and need to support the development of RDF Technologies.

Han et al.[2000][8]gives XML structure for maintaining multi-gadget Web Searching. The system demonstrates a XML-construct metadata arrangement record depend with respect to the CC/PP convention to permit clients to distinguish the entrance benefit bunches. With Web Splitter, all Web pages are actualized as XML records by pre-characterized labels giving mappings to the identical access benefits. Subsequent XSL templates are included to gadgets to change over the tweaked XML to the suitable gadget understandable dialects. MyXML is a XML/SL based format motor to determine the matter of gadget sway for totally isolating the substance since its design data. Identified with the WebSplitter, MyXML utilizes the XML/XSL blend to comprehend gadget freedom. In any case, MyXML characterizes a complete arrangement of language structure essentials that necessities expectation to learn and adapt generally procedure won't return in great results.

Cardoso[2002][9] proposed a model which concentrated on Semantic Web Services to find web benefits and give the interoperability among them despite the heterogeneity normal for web administration. It doesn't address the issue that a web administration can give distinctive QoS levels and by what means can client chooses which level ought to be utilized and

additionally how to accomplish the best end to end QoS in web administration work processes.

E.Michael Maximilien et al.[2002][10] brought a reference model named Web Service Agent Proxy (WSAP) keeping in mind the end goal to choose web administrations which meets the QoS necessities in writing, they portrayed the notoriety idea model of web administration in point of interest so that the notoriety data can be sorted out and be shared by pertinent performing artists. In any case, how to get notoriety score is not determined in these literary works.

P. Farkas et al.[2003][10] gives QoS representative amongst customers and administration suppliers to find the QoS mindful web administrations in UDDI. For this, they proposed programming design to give QoS empowered web administrations. In any case, the issue in this methodology is that no point by point data about QoS intermediary is given, for example, how it is composed and how it is worked.

A.ShaikAli et al.[2003][11] proposed a methodology which depends on expansion of UDDI business administration structure, however potential QoS changes are not considered.

P. Octavian et al.[2003][12] proposed the method based on k-d tree for indexing of data. In proposed work it has been utilised for good execution as well as designing of requests at data centres.

S.Kalepu et al.[2003][13] proposed the new QoS attributes to verify the system i.e. reputation. Reputation has been applied to measure the average rating of web services.

Hongan Chen et al.[2003][14] displayed a depiction and an execution of merchant based design for controlling QoS of web administrations. The specialist goes about as a middle person outsider to make web administration's determination and QoS transaction in the interest of the customer. Designation of determination and transaction raises dependability issues fundamentally for customers. Execution of the representative is not considered in this methodology. In addition, execution of the representative can be a key to the achievement of any proposed design. In the event that the client does not get a reaction to his/her solicitation with a worthy reaction time, he/she will change to another supplier. Presently, with an expanding number of Web administrations giving comparable functionalities, more accentuation is being set on the most proficient method to discover the administration can do), and non useful necessities which are useful prerequisites (what the administration can do), and non useful necessities, for example, the cost and nature of administration ensured by an administration supplier in **Gouscos [2003] [15].** UDDI registries are driven for WS disclosure. They help web administrations life cycle administration, yet UDDI registries are little, private and has no power over web administrations distributed data. They just backing syntactic

recovery and give no backing to quality mindful administrations. Proposed another structure for more brilliant WS disclosure that furnishes customers with QoS data which improve the determination procedure and lessen the disappointment chances by getting supports or proposals from different administrations or specialists about every administration. In this paper, Gouscos proposes another way to deal with element Web administrations revelation that models real traits, for example, QoS and cost. The model can be executed utilizing fundamental measures, for example, WSDL. The upside of this model is its effortlessness and simple execution over existing norms, yet at the same time the QoS measuring is subjective and client driven.

E.M Maximilien et al.[2004][16] use current web administration advancements, for example, WSDL and UDDI, which are for distributed and finding web administrations, however consider just client usefulness prerequisites and bolster plan time, or static administration revelation. Non utilitarian prerequisites, for example, QoS, are not bolstered by current UDDI registries.

Ran [2004][17] proposed an augmented model containing the customary parts: administration supplier, administration purchaser and UDDI registry, alongside another segment called a Certifier which confirms the QoS of a web administration before its enlistment. In spite of the fact that this model joins QoS into the UDDI, it doesn't incorporate shopper criticism into the administration revelation process. In any case, it needs bolster for the dynamism of web administrations.

Kalepu, Krishnaswamy and Loke [2004][18] evaluated the notoriety of an administration as an element of three variables: appraisals made by clients, administration quality consistence, and the progressions of administration quality conformance after some time. In any case, these arrangements don't consider the unwavering quality of QoS reports created by clients, which is imperative to ensure the precision of the QoS-based Web Service choice and positioning results.

Zeng et al.[2004][19] has talked about a worldwide arranging way to deal with select composite Web administrations. They have proposed a peaceful straightforward QoS model containing five QoS attributes, to be specific value, term, accessibility, unwavering quality, and notoriety. They use straight programming to take care of the issue of determination advancement in administration piece, however with regards to run-time choices, this methodology is excessively perplexing.

Yucheng Liu et al.[2004][20] proposed a method that could adapt changing environmental conditions which shown the improvement in terms of accuracy but it was less practical.

Chenliang Sun et al.[2004][21] has analysed two replication strategies and number of implementation substitutions for UDDI replications. In this research, the system has been tested on more replicas and focuses on response and execution times. The clients connect to UDDI registry serially by using save_business method and reads the three attributes of unique table. The modifications took place later for performing the business details, adding descriptions, discovering URL's and the contacts. It is being concluded that an aspect oriented approach has better mechanisms for integrating the replications with the business logics. This paper does not attempt to invent yet another strategy.

Werner et al.[2004][22] proposed a method that could adapt changing environmental conditions and also overcomes the individual difference of evaluation. The main problem is in the complexity of service selection mechanism. Although the above mentioned methods have shown the improvement, but they may be less practical when some additional information or the personalized knowledge made unavailable.

Jia Zhang et al.[2005][23] presented a service oriented multimedia componentization model for supporting QOS (Quality of Service).The enhancement is projected to make possible Simple Object Access Protocol (SOAP) and Composite Capability/Preference Profiles (CC/PP) protocols to recover their flexibility to implement multimedia web services. The application which is being selected for implementation is a distance-learning environment. The designing is such that every client can develop the threads and then can communicate with the server by requesting the multimedia services. The tests are being performed on the server machine by MWSS agent and without MWSS agent. The problem that arises in proposed work is that it takes longer because of the extra operation time spent on the MWSAgent.

Serhani et al.[2005][24]shows web administration engineering which utilizes a stretched out UDDI registry to bolster administration choice in light of QoS, yet just the accreditation methodology is utilized to confirm QoS and no data is given about the QoS Specification. He proposed ranking algorithm based on QoS which produced third party between client and provider but no information is specified about the QoS Specification

Le Hung [2005][25] presented QoS-Based Service Selection and Ranking mechanism in which Web Services Selection Criteria was considered using SOAP Protocol but Ranking criteria was not considered.

R. Panigrahy et al.[2006][26] proposed the method for text categorization based on the usage of the data mining i.e. k- d tree and neighbors method.

Chia-Hui et al.[2006][27] discussed clustering technique k-mean which is partition-based clustering method. K-mean firstly initialize the centre and then calculates the distance of

another element k-mean creating the cluster of those elements whose distance is very less from the centre. K-mean is very simple, flexible and easy to understand and implements. Kmean works for numeric dataset .Author need to improve the efficiency of the k-mean clustering algorithm and also to reduce the computation time taken by it.

L.Hung et al.[2006][28] proposed the information extracting using two algorithms called Bloom filters and matching information. In proposed work, ranks can be utilised. From Robustness of method has been guaranteed from result simulation.

Ziqiang Xu et al.[2007][29] provided a reputation based Web Service Discovery Model that contains an extended UDDI to accommodate the QoS information but they did not provide any verification process for QoS.

Jing Ni et al.[2007][30] has illustrated a Semantic Web e-business through which the information sources and administrations are made accessible by utilizing SWS, outlined by ontologies, allowing interoperability and thinking to produce a complete reaction customized to client objectives. The proposed module is partitioned into three layers. The client communication has upheld the client in distinguishing and gathering the data for administration usage. This work utilized E-Business show and made an endeavour for incorporating the advancement of Semantic Web Services with the utilization of WSMO into ISR-III structure. The fundamental issue of this technique is that no unequivocal semantic data is typically characterized, robotized understanding of the WSDL portrayal is restricted to situations where the supplier and requester expect pre-concurred ontologies, conventions and shared information about operations k-means, nor k-mediods work on very large scale data.

M. Jose et al.[2007][31] has proposed the method that uses the information extraction techniques and some transformations are being made in automating the larger amount for extraction of resource process. The information retrieval metrics are used for evaluating the extraction. This research provides a traceable process for extracting hierarchal model for generating large real life applications.

Liu et al.[2007][32] proposed data mining method for predict the requests of customers. In this paper the customer sessions will be recorded using n- gam method.

Yang et al.[2007][33] proposed the usage of the hyperlinks for analysing the pages of the packages. This research provides a traceable process for extracting hierarchal model for generating a large real life applications.

Y.Liu et al.[2007][34] proposed the solution of embedded dynamic web based on PHP, HTML and apache. The simulation results shows that proposed method provides the high security level and better adaptability.

H.Alani et.al[2008][35] proposed the semantic technology for sharing of large number of applications while getting benefits of reuse.

M. Sehrani et.al[2008][36] proposed the architecture of the web service selection method. based on brokers cooperation. Each broker manages the services by its domain. In proposed work three broker selection has been done like random, round robin and cooperative brokers.

N. Olfa et al.[2008][37] proposed the framework for completing the web designing outlines based on web records. It provides the tools for customer profiles and external data usage n website.

T. Rajendran and Dr. P. Balasubramanie [2009][38] has exhibited an investigation and investigation of Web administrations revelation by utilizing QoS Management frameworks. The primary reason for web administration disclosure is to utilize ideal web administrations for a predefined errand. For assessing and positioning of the hopeful web administrations,QoS assumed an imperative part which gives the usefulness which is albeit anticipated. A web administration display that gives web administrations for giving the capacities and the business operations that can be further conveyed by the web is utilized. The whole gang of this exploration concentrated on the methodologies and the calculations for coordinating and positioning of QOS online administrations. In any case, the greater part of these methodologies don't give ensures with regards to the precision of the QoS values after some time or having avant-garde QoS data.

Dou et al.[2009][39] presented the web mining structure and its three classes like web usage, substance mining and structure mining. First class contains the data gathering process, second class contains the web records and third class contains the web customization.

Pak and Alexander et al.[2010][40] has been shown how to consequently acquire corpus for mining of huge information and assumption examination purposes. At that point it performs standard semantic examination system of the gathered corpus and clarifies procured wonders. By utilizing the corpus, author fabricates a supposition classifier, which is equipped for deciding positive, unbiased and negative notions for the entire record. Test results demonstrate that the proposed strategies are more productive and perform better when contrasted with beforehand proposed procedures. Creator worked with English; in any case, this proposed method can likewise be utilized with other dialect.

Sujatha et al.[2010][41] proposed the characterization of records written in Arabic dialect utilizing neural system (ANN). This procedure has been utilized as a part of constrained adaptation in earlier years. Arabic records have been gathered from the Arabic content corpus. Firstly, preparing and afterward testing have been done utilizing ANN classifier. Notwithstanding this, the weighted plan for highlight extraction has likewise been utilized.

35

Agarwal and Apoorv [2011][42] clarified one such prominent miniaturized scale blog named as Twitter and construct models to arrange the "tweets" into positive and negative supposition or they can be impartial. Creator construct novel models for two classification: initial one is a double errand of ordering slant of clients into positive and negative classes and second is a 3path undertaking of characterizing estimation of clients into positive, negative and nonpartisan.

Netra Patil et al.[2011][43] presented that the web service is the set of methods that can be accessed over internet. For doing this, one person requires the selection of service. Web services are important for various industrial applications. The author introduced tModel in order to solve the limitations in UDDI registry. But still due to category bag of tModel, problems exist.

L. Vijayashri et al.[2012][44], a complete pre-handling methodology is being proposed to pre-process the web log for extraction of customer cases. Data cleaning computation removes the irrelevant areas from web log and isolating figuring hurls the uninterested properties from log record. Customer and sessions are perceived.

B. Geeta et al.[2012][45] proposed the usage of the hyperlinks for analysing the pages of the packages. This research provides a traceable process for extracting hierarchal model for generating a large real life applications.

P. Nithya et al.[2012][46] communicated that the World Wide Web keeping up its change at an inconceivable pace. The information open in the WWW is a gateway and a most of the way to complete business. Web mining is the extraction of empowering and helpful realities and inalienable information from antiquated rarities or exercises related to the WWW. Web Use Mining (WUM) puts a push to center critical information from the discretionary data obtained from the correspondences of the customer with the Web. WUM has wound up being an amazingly basic for productive Web website page association, creating versatile web locales, business and upkeep administrations, personalization and system movement stream examination et cetera. WUM includes three phases, specification.

Jebaseeli and A. Nisha et. al[2012][47] presented the primary goal of Opinion Mining is to decide the extremity of remarks (positive, negative or nonpartisan) by separating elements and parts of the article that have been remarked on in every report. Ponders identified with supposition mining of huge information, on the ramifications of monetary effect because of the audits, issues about break of security are given consideration.

Shariq Hussain et al.[2013][48] proposed the three prominent open-source instruments and look at them as far as elements, ease of use, execution, and programming prerequisites. The after-effects of the examination will help in reception and utilization of these instruments
furthermore advance improvement and use of open-source web administration testing apparatuses. Correlation of various testing devices is an unpredictable assignment because of the way that testing devices may not agree to same test criteria, i.e. one device may be able to test throughput (for our situation JMeter and soapUI), while another apparatus i.e. Tempest, does not have this criteria. Moreover, one device may have better execution in one experiment, while poorer in other test criteria.

Ram Mohan Reddy et al.[2013][49] surveyed the writing accessible for the quality traits of execution, adaptability and distinguished the issues that influence the quality ascribes identified with Web Services. This work proposes reasonable trust estimation for technique for recently web benefits yet not for existing ones but to characterize ideal method for arranging these web administrations.

Anuradha et al.[2013][50] gave that the expedient development of online data there was expanding need for devices which are exceptionally useful in dealing with the high dimensional information and finding sifting. Content order was a kind of administered learning undertaking whose fundamental target was to classify report records into the predefined gatherings or classifications. Stages which are ensnared in content grouping were of stemming, gathering information set, pre-preparing, alongside executing the classifier and also their execution measure. There are various learning methods which are being used for Text arrangement which are Support Vector Machine, Naïve Bayes, k-closest neighbour choice tree, Back Propagation NN and so forth. This paper demonstrates the consequence of content grouping through using Back Propagation NN and pertinence component (rf) as a term measuring technique.

O. Mohammad et al.[2013][51] proposed KNN method based clustering to find the k-closet neighbour based on spatial grouping. Proposed KD tree worked on 2D and 3D space having database of n focuses and then it finds the k- closet neighbour.

Balahur and Alexandra[2013][52] distinguished three subtasks that should be tended to: characterizing the objective; isolating the awful and uplifting news content from the terrible and great conclusion communicated; lastly investigation of unmistakably said feeling that is communicated unambiguously, not requiring understanding or the use of world learning. Besides, we recognize three unique perspectives on daily paper articles (content, writer and per user), which must be handle contrastingly while examining supposition.

Gupta and et al.[2013][53] presented that social media is very important for sentiment analysis. It aims to provide the attitude of the speaker in a document.

Dongjin et al.[2014][54] gives XML structure for maintaining multi-gadget web searching. The system demonstrates a XML-construct metadata arrangement record depend with respect to the CC/PP convention to permit clients to distinguish the entrance benefit bunches. With Web Splitter, all Web pages are actualized as XML records by pre-characterized labels giving mappings to the identical access benefits. Subsequent XSL templates are included gadgets in order to change over the tweaked XML to the suitable gadget understandable dialects.

Dev. Deepika et al.[2014][55] proposed algorithm that uses standard deviation that reduces the total time to formulate the cluster by simple k-d tree method. The proposed method divides the square root distance with standard deviation. This k-d tree perform much better than that of traditional methods.

Ramakrishna M. et al.[2014][56] discussed the modified version of k-mean algorithm for formulating clusters. Modified k-means algorithms can perform clustering very effectively on Categorical data sets. The modified k-mean use pruning methods by this number of iterations that are needed for calculating distance are reduced. This method is well suited for categorical data sets and also improving the efficiency of the k-mean clustering algorithm and reduces the overall computational time. K-mean is efficient for small dataset only and if noise is present it cannot work efficiently.

Chengying Mao et al.[2015][57] given a search based prediction framework for addressing the QOS ranking problem. The PSO (Particle Swarm Optimization) algorithm is being implemented for optimizing the order of services with the QOS records. For filtering the neighbouring as per the user, the similarity among the users with the occurrence probability of the service pairs is presented and for evaluation of the service ranking, the top k-neighbours are used by means of supporting QOS information. High quality ranking results are obtained by the use of the fitness function and experimented with the real world QOS data.

Muhammad Waqas Khan and Eram Abbasi[2015][58] have focused on the REST (RESTful) and SOAP (Simple Object Access Protocol) web based services. The SOAP architecture is implemented in the XML encoded message which is transmitted on the HTTP (Hyper text transfer protocol). The WSDL file is written first on the server side with the XML schema types and later the WSDL is made public for reaching the particular public server by using SOAP enabled web based server. While in the case of REST, the HTTP client library is used for interacting with the REST server instead of the SOAP client.

Zainab Aljazzaf [2015][59] has shown a QOS bootstrapping solution for the web services and a QOS bootstrapping framework is built.SOA (Service Oriented Approach) work is being extended. A prototype is implemented for supporting QOS bootstrapping. QOS is used as a factor among the services and shows better performance in service delivery, selection and composition. There is a need to extend the work to bootstrap other QoS in the QoS model. **Michael Athanasopoulos and Kostas Kontogiannis [2015][60]** has proposed the method that uses the information extraction techniques and some transformations are being made in automating the larger amount for extraction of resource process. The information retrieval metrics are used for evaluating the extraction. This research provides a traceable process for extracting hierarchal model for generating a RESTer API for large real life applications.

Horakova and Marketa [2015][61] presented a model which gathers tweets from long range interpersonal communication destinations and in this manner give a perspective of business knowledge. In this structure, there are two layers in the feeling examination instrument, the information handling layer and opinion investigation layer. Information handling layer manages information accumulation and information mining, while feeling examination layer utilize an application to introduce the after-effect of information mining.

2.2 INFERENCES DRAWN OUT OF LITERATURE SURVEY

Recently, Web Service has become an important issue for developers. Selecting a specific service is a crucial task. Some approaches develop extensive description and publication mechanisms while others use syntactic, semantic, and structural reviews of Web service specifications. It is very crucial for finding the most suitable web service from a large collection of web services for successful execution of applications. In many cases, the value of a QoS property may not be precisely defined. Recently, Fuzzy is considered as the dominant approaches in Web services which can deal with fuzzy constraints have been proposed. Therefore, Fuzzy Logic can be applied to support for representing such imprecise QoS constraints. In this work, we will present an overview which focus on developing fuzzy-based approach along with neural network, LDA and SVM for Web service discovery as well as Naïve Bayes classifier will be imposed for QoS factors.

2.3 ISSUES IN LITERATURE REVIEW

Issues	Literature Review
Protocol Comparison	SOAP
Registry Issues	UDDI,WSDL
Security Issue	XML
Model Issue	-Non Adaptable -Project Dedicated Model
Optimization Issue	Less Optimized

Following Issues has been raised from literature review:

- 1. **Protocol Comparison**: Code used in SOAP protocol is very plumbing. Size of XML packet becomes large.Everytime while communication among protocols, there is a need to compress XML packets.
- 2. **Registry Issue**: UDDI supports only syntactic information in WSDL format. It does not provide semantic details.
- 3. **Security Issue**: In existing XML based tModel, the entire category bag needs to be transferred while communication among protocols. Due to this reason, Security becomes reduced.
- 4. **Model Issue**: Existing Models which studied during literature survey are suitable for particular applications and dynamic changes are not possible.
- 5. **Optimization Issue**: Optimization Issue arises due to ignorance of irrelevant data by many researchers which in turns reduce accuracy.

CHAPTER-3: PROPOSED WORK AND METHODOLOGY

3.1 PROBLEM STATEMENT AND OBJECTIVES

Problem for proposed work is stated as "Prioritization of Quality factors in Agent Based Web Services" by utilizing Naive Bayes Classifier, Neural Network, Fuzzy logic, LDA and SVM.

RESEARCH OBJECTIVES

• Focus on Non-Functional properties of web services.

Non-Functional properties means how the service can performs .These are Throughput, Response Time, Content availability and Accuracy.

• To investigate how dynamic web service discovery can be achieved.

In Dynamic Web Service Discovery, construction and many other things are controlled by server side script. Any layman can know only about the behaviour of any web service but to know about hidden or background detail of any web service at run time is impossible.

• Focus on quantifiable measurements, prioritization of web services & authenticity of QoS information.

- For quantifiable measurements, calculations of quality scores of every web service like throughput and response time will be quantified.
- After analysed quality score of similar web services, ranking will be provided to every web service in order to distinguish among similar web services.
- For the purpose of Authenticity, irrelevant records from the relevant ones will be removed. Various accuracy measurements will be performed in terms of Recall Rate and Precision Rate.

To achieve these goals various sub-objectives has to be followed up as shown below;

- 1. To measure QoS (Quality of Service) based on web services.
- 2. To propose a web service model based on various algorithms i.e. Naïve based approach, Fuzzy Logic, Neural Network, LDA and SVM method.
- To check the performance of proposed work using QoS Parameters like Recall rate, Precision rate, Response Time and Throughput.

3.2 PROPOSED WORK

QoS selection has become essential for real time applications. Number of web applications requires guaranteed services processing. There are number of standards that support the web services that are SOAP (Simple Object Access Protocol), WSDL (Web services description language), UDDI (Universal Description Discovery Integration), and XML (Extended Mark-up Language). The web services that are available these days do not support Quality of Services (QoS) or some of the non-functional aspects. Quality of Service with the service consumer is necessary for adopting the best quality web services. The existing approaches that have shown dissimilarity in the results of the same web services are in number.



Figure 3.1 Web Service Model

Above Figure 3.1 shown is the base model (Web Service Model) that has the service provider that selects service by using the service provider. Service provider publishes the description of the service functions. Service consumer faced most problems in an open environment when the same function is offered by the cluster of services. XML (Extended Mark-up Language) and SOAP (Simple Object Access Protocol) protocols are used for the communication between Service Provider and Consumer. It is necessary that a service consumer should have knowledge about the working of the web services. Only syntactic information is provided by the UDDI (Universal Description Discovery Integration). But Semantic information is mandatory as well for the consumers that contains the response time, availability, content, throughput and many more and if the solution of this problem is provided, then one more problem can occur, i.e. how in UDDI the Quality of service information on same web services will be published by the service providers. The need of updating and saving information can occur in UDDI after publishing of QoS.



Figure 3.2: Proposed Model

For storing the web services, number of frameworks and architecture used UDDI registry. T-Model (Technical model) was represented for solving the problem of UDDI that would store QoS values for every web service. But there are number of problems that need to resolve yet. The UDDI was not proposed for the web service discovery in the search engines. This can make passive easily because of the voluntary registration. It doesn't have any guarantee for the information contains. Between UDDI and Current Web, a disconnection exists and UDDI has not the capability to provide Quality of Service for the web services that are registered.

For those customers who use to evaluate and compare the quality factors of a particular thing for selecting the best one based on QoS requirements from different shopping sites, this model is basically helpful. Different websites produce similar results will differ according to their QoS factors. For example: A customer wants to buy a mobile cover and requests through the search engine. The Broker receives the request and interacts with the service provider and checks whether the response exists or not in UDDI registry. If the customer makes number of requests, then with the number of responses it will return to the client. The objective of the research is to find out the best response, according to the QoS factor. By adding the functionalities to search agent, the performance of the broker has been enhanced in this model. From database to search a particular query requested by the customer is the task of the customer. To check the existence of results in the registry, the search agent interacts with the service provider. To search a particular request, fetch and display of multiple results in front of the user is limited role of the search agent. The efficiency of the search will increase by adding other functionalities. All the web pages return similar results linked to a website are fetched by the web crawler under Search Engine Optimization. Then indexing of all the fetched web pages is done and then placed in a large database and later from there it can be retrieved. To identify the words and expressions that describe and assigns the page to a particular keyword the process of indexing is used. All the web pages are being forwarded by the service provider equivalent to the user demand to quality manager for finding the best quality factors. Later the results which are calculated by the priority means through the service provider are being forwarded to the broker with quality manager and to the same client in the end. As shown in Fig.3.3, we have utilized basically ranking algorithms like Naïve Bayes, LDA, Fuzzy logic, Neural Network and SVM to rank the web services. The work investigates the proposed model on different parameters i.e. throughput, response time, precision rate and recall rate in MATLAB 2010a.

3.3METHODOLOGY



Figure 3.3: Workflow of Proposed Model

3.4 TOOLS USED IN METHODOLOGY

Computer	Core 2 Duo or higher
RAM	3 GB
Platform	Windows 7
Other hardware	Keyboard, mouse
Software	PHP 5,Symfony Framework,MySQL,Matlab 7.0.4

Table 3.1: Tools used

3.4.1 Symfony Framework

Symfony Framework is the framework in PHP for developing websites.

3.4.1.1 Installation

Symfony Framework uses the composer to manage the applications of Symfony.





Symfony Profiler				search on s	ymfony.com	Search
http://symfony.app/weld Method: GET HTTP Status: 200	COME IP: 127.0.0.1 Profiled	d on: Fri, 18 Mar 2016 1	2:36:29 +0100 Token: da4f73			
Last 10 Latest O Search	Performance I	metrics				
O Request / Response	37 ms	12 ms	3.20 мв			
C Performance	Total execution time	Symfony initialization	Peak memory usage			
🖹 Forms						
😭 Exception	Execution time	eline				
🗓 Logs	Threshold 2	ms (timeline only d	isplays events with a duration longer than this thre	eshold)		
🛠 Events	default section	event_listener eve	nt_listener_loading 📕 template 📕 doctrine 📕 pro	pel		
■ Routing	kernel.request 8 ms / 2.2	2 MB				
2 Security	Symfony\Comp	onent\Security\Http\Fire	3wall 6 ms / 2.2 MB			
🔳 Twig			controller 15 ms / 3.2 MB			
Doctrine				kernel.response 8 ms	/ 4 MB	
📼 F-Mails				Symfony\Component\HttpKernel\Eve	entListener\ProfilerListen	ier 8 ms / 4.2 N
					kernel.termina	.te 5 ms / 4.2 №
			Symfony	\Bundle\SwiftmailerBundle\EventListene	r\EmailSenderListener	1 ms / 4.2 MB
Configuration						

Figure 3.5 : Symfony Framework Performance Metrics

In above Figure 3.4 and Figure 3.5, shows the starting view of Symfony Framework 3.0 and Shows the performance metrics of logs.

3.4.1.2 Twitter Bootstrap

For designing of the application, twitter bootstrap has been utilised. There are various other components for designing but bootstrap is very essential. It provides tables, buttons, and forms in easy way. Moreover module setup is very easy.

3.4.2 MATLAB

MATLAB is the acronym for MATRIX LABORATORY. It works on metrics and arrays. It is very useful in mathematical research areas and education too. MATLAB has powerful tools that work in 2D as well as 3D. There are plenty of functions in MATLAB and some of the commonly used windows are: command window, workspace, help menu, history window etc. as shown in Figure 3.6 to Figure 3.14.





MAILAB 7.10.0 (R2010a)	and the second se			
File Edit Debug Parallel Desktop Window	w Help			
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Current Folder: C:\Users\sayal\Documents\MATLAB			
Shortcuts 🔄 How to Add 🖪 What's New				
Current Folder 🌐 🏕 🗙	Command Window	X 5 D	Workspace	× 5 ⊡ !+
🔹 🔹 🕌 « MATLAB 🔹 🔎 🌞-	(1) New to MATLAB? Watch this <u>Video</u> , see <u>Demos</u> , or read <u>Getting Started</u> .	×	1 d 1 i i i i i i i i i i i i i i i i i	Select data to 👻
Name A	MATLAB desktop keyboard shortcuts, such as Ctrl+S, are now customizable. In addition, many keyboard shortcuts have changed for improved consistency across the desktop. To customize keyboard shortcuts, use <u>Preferences</u> . From there, you can also restore previous default settings by selecting "R2009A Windows Default Set" from the active settings drop-down list. For more information, see <u>Help</u> . <u>Click here</u> if you do not want to see this message again. /s >>		Name ▲ "" Command History =6/6/14_11 =6/6/14_11	Value Value 01555 AM5 1104 AM5 9:43 AM5
Details 🗸				9:48 AM%
Select a file to view details				9:27 PM* 9:31 PM* 8:32 PM* 2:27 PM* 9:06 PM* 3:26 PM* 3:26 PM* 5:40 PM*
Carral Decidi			L	OVP

Figure 3.7 Command History



Figure 3.8 Workspace



Figure 3.9 Current Folder





Create New GUI Open Existing	GUI
GUIDE templates	Preview
 Blank GUI (Default) GUI with Uicontrols GUI with Axes and Menu Modal Question Dialog 	BLANK
Save new figure as: E:\SM	MARTTECH\Matlab Program\untitled1.f

Figure 3.11 NEW GUI

4	GUIDE Quick Start – 🗖 🗙			
Create New GUI Open Existing GUIDE templates Blank GUI (Default) GUI with Uicontrols GUI with Axes and Menu Modal Question Dialog	GUIDE Quick Start – GUI Preview BLANK			
Save new figure as: C:\Users\KARANBIR\Documents\MATLAB\L Browse				
	OK Cancel Help			

Figure 3.12 Other Options for GUI



Figure 3.13 To Draw a GUI



Figure 3.14 Editor Window

3.5 CLASSIFICATION ALGORITHMS

3.5.1 Fuzzy Model

Fuzzy Model is the generalized model of previous classic models. As the output is not limited to only 0 and 1, so the theory of fuzzy logic is introduced. It is also known as diffuse logic. Difference between fuzzy logic and classical model is introduced using membership functions. Consider a finite set [62],

$$C = \{c1, c2, c3..., cn\}$$
 (3.1)

It is the universal set. Now according to graphical representation, suppose fuzzy sets has only two elements c1 and c2. So the degree of fuzzification can be called as entropy. Therefore entropy can be shown as:

$$E=\frac{f_1}{f_2}$$
; Where f1 and f2 are the distances (3.2)

Procedures on Fuzzy Set

The intersection of D and M is defined as

$$(D \cap M)(u) = \min\{D(u), M(u)\} = D(u) \wedge M(u),$$

For all $u \in V$.

The union of D and M is defined as $(D \cup M)(u)$ = $max\{D(u), M(u)\} = D(u) \lor M(u)$, for all $u \in V$.

- The complement of Fuzzy set is fuzzy set S is defined as $(\neg D)(u) = 1 D(u)$.
- ➤ Law of excluded middle can be shown as: Let $D(u) = 1/2, \forall u \in Y$, then it is easy to see that $(\neg D \lor D)(u) = max \{\neg D(u), D(u)\}$.

Law of Contradiction can be shown $\operatorname{Let} D(u) = 1/2, \forall u \in Y$, then it is easy to see that $(\neg D \land D)(u) = \min \{\neg D(u), D(u)\}.$



Figure 3.15 Fuzzy logic in MATLAB

The toolbox of fuzzy logic allows to run the stand-alone C programs directly. This is made possible by a stand-alone Fuzzy Inference Engine that reads the fuzzy systems saved from a MATLAB session.

3.5.2 Support Vector Machine (SVM)

SVM is proposed by Vapnik to develop a classification algorithm. Various features of SVM include: Less training time, easy to implement, high accuracy and robust [63].

Support vector machines (SVMs) have the hyperplane that classifies the various variables as shown in Figure 3.16:



Figure 3.16 Support Vector Machine

Equation of hyperplane can be written as below:

$$\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = \mathbf{0} \tag{3.3}$$

Proposed hyperplane gives the function $f(c) = sign(e \cdot c1 + n1)$

Which correctly classifies the training data (and hopefully other "testing" data it hasn't seen yet). Also the hyper canonical plane ca be described below. That is, we consider those that satisfy:

$$ci \cdot e + n \ge +1 \text{ when } ui = +1 \tag{3.4}$$

$$ci \cdot e + n \le -1 \text{ when } ui = -1 \tag{3.5}$$

or more compactly:

$$ui(ci \cdot e + n) \ge 1 \, \forall i \tag{3.6}$$

3.5.2.1 Basic operations utilize the SVM

Multiplication by scalar

Consider a vector $\mathbf{a} = \{a_1, a_2, \dots, a_n\}$ and a scalar c and it can be shown as

$$\{\operatorname{ca}_{1},\operatorname{ca}_{2}\ldots\ldots\ldots\operatorname{ca}_{n}\}$$
(3.7)

Addition of Vectors

Consider a vector $\mathbf{a} = \{a_1, a_2, \dots, a_n\}$ and $\mathbf{b} = \{b_1, b_2, \dots, b_n\}$, then addition ca be shown as below:

$$\{a_1 + b_1, a_2 + b_2 \dots \dots a_n + b_n\}$$
(3.8)

Subtraction of Vectors

Consider a vector $\mathbf{a} = \{ \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \}$ and $\mathbf{b} = \{ \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b} \}$, then subtraction ca be shown as below:

$$\{a_1 - b_1, a_2 - b_2 \dots \dots a_n - b_n\}$$
 (3.9)

Euclidean Distance

Consider a vector $\mathbf{a} = \{ \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \}$

Then it can be write as:

$$||a|| = \sqrt{a_1^2 + a_n^2} \tag{3.10}$$

Dot Product

Consider a vector $\mathbf{a} = \{a_1, a_2, \dots, a_n\}$ and $\mathbf{b} = \{b_1, b_2, \dots, b_n\}$, then dot product can be shown as below:

$$\{a_1, b_1, a_2, b_2, \dots, \dots, a_n, b_n\}$$
 (3.11)

3.5.3 Neural Network

Research in the field of neural networks has been growing from past very years. Neural Network structure is similar to human brain consists of small neurons. The Neural Networks are developed to work like human brain to reduce the computational tasks. Basic elements of Neural Network in MATLAB are epochs, gradient, validation step as shown below in Fig. 3.17 and Fig.3.18. There are three types of layers i.e. input layer, hidden layer and output layer as shown below in mathematical model. The structure of neuron can be shown as below :



Figure 3.17 Neural Network

The net input can be shown as:

$$n1^{k1+1}(i) = \sum_{j=1}^{s1k1} w1^{k1+1}(i,j)a1^{k1}(j) + b1^{k1+1}(i)$$
(3.12)

The unit i is given by

$$a1^{k1+1}(i) = f1^{k1+1}(n1^{k1+1}(i))$$
(3.13)

This recurrence relation is executed at the final layer.

$$-F1^{M1}(n1^{M1})(t1_{q1} - a1_{q1})$$
(2.14)

Neural Network				
Hidden Lay	er Output Layer			
Input W +		Output		
Algorithms				
Training: Scaled Conjugate Gradient (trainscg) Performance: Mean Squared Error (mse) Data Division: Random (dividerand)				
Progress				
Epoch: 0	15 iterations	1000		
Time:	0:00:00			
Performance: 0.440	0.0290	0.00		
Gradient: 1.00	0.0121	1.00e-06		
Validation Checks: 0	6	6		
Plots				
Performance	(plotperform)			
Training State (plottrainstate)				
Confusion (plotconfusion)				
Reciever Operating Characteristic (plotroc)				
Plot Interval: U				
Validation stop.				
	Stop Training	Cancel		

Figure 3.18 Neural Network Model

These networks have been trained in the direction of performing complex functions in numerous areas, which also encompasses speech, pattern recognition, vision, control systems, identification, and classification. Neural Networks could also be trained towards resolving issues which are challenging for conventional computers or human beings. Even though there are numerous different types of learning rules utilized via Neural Networks, this specific demonstration is concerned simply with one and only the delta rule. This specific delta rule is every so often used by the utmost common class of Artificial Neural Networks entitled as Feed Forward Neural Networks'. The training criteria of NN can be summarized below:

- 1. Input is given to input neurons.
- 2. Obtained output response is compared to input data.
- 3. Error data is utilised to manage the weights attached to neurons.
- 4. Hidden units find outs its error during back signal.
- 5. Then weights gets updated in the end.

There are two types of ANN, based on the architecture of connections:

- 1. BPNN
- 2. FFBPNN

In BPNN delta rule consists of two phases.

- 1. In first phase compute the output values.
- 2. Input value in computed and it is compared to output value.
- 3. Backward pass is doe to get the weight measurement. Below figure shows the working of BPP.



Figure 3.19 BPNN Process Flowchart

The algorithm is described below:

- 1. Initialise weights
- 2. Initialize inputs and each input unit is represented as (Ci, $i = 1 \dots m$).

3. Each hidden unit is represented as (Xj, j = 1, ..., a) and weighted input Signals are represented below,

$$X_{in} = B_{oj} + \sum_{i=1}^{m} C_i B_{ij}$$

Where B0j: hidden unit j bias.

Bij: Weight between output and input unit.

4. Each output unit gets its output as show below;

$$Uim_k = E_{0k} + \sum_{j=1}^{a} X_j E_{jk}$$

5. Output unit having activation function is shown belwo:

$$U_k = g(uim_k)$$

- 6. Finding of back propagation error.
- 7. Each output unit gets its output by changing error values.

$$\delta_k = (\mathbf{y}_k - \mathbf{l}_k) \mathbf{g} \; (\mathbf{uim}_k \;)$$

8. Calculation of weight function.

9. Calculation of bias correction term

$\Delta E_{ok} = \partial \delta_k$

- 10. Update weights and biases:
- 11. Each output unit (Yk, k = 1, ..., m) updates its bias and weights (j = ... p) :

Ejk(new1)=Ejk(old1)+ΔEjk

12. Test stopping condition.

3.5.4 Naive Bayes Classifier

It is the classification method that is basically based on bay's theorem. It assumes that the features which are not included into the same class will be reduced out. It is very easy to build and apply. It can be written as probability A(v|c) from A(v), A(v) and A(c|v). Look at the equation below in Figure 3.20:



Figure 3.20 Naïve Bayes Classifier

3.5.5 LDA

Linear Discriminant Analysis is utmost commonly utilized as dimensionality lessening method in the pre-processing stage for machine learning applications in addition to designclassification. The main objective is to project a specific dataset on top of a lower-dimensional space using virtuous class reparability so as to decrease computational prices as well as also evade overfitting. The novel linear discriminant was first designated for a two-class issue, in addition it was then afterwards widespread as "Multiple Discriminant Analysis" or "multiclass LDA" through C. R. Rao in the year of 1948. Linear Discriminant Analysis is "controlled" as well as calculates the guidelines ("linear discriminants") which would probably signify the axes that are applied to make the most of the separation amongst multiple type of classes.

Below are the five basic steps as shown in Figure 3.21 utilized for implementing a LDA technique;

- Step 1 : Calculate the d-dimensional mean vectors intended for the dissimilar classes from the specific dataset.
- Step 2 : Calculate the disseminate matrices i.e. between-class as well as within-class scatter matrix.
- **Step 3 :** Evaluate the eigen vectors (e1, e2... ed) as well as corresponding eigen values (λ 1, λ 2, ..., λ d) for the disseminate matrices.

- Step 4 : Sorts the eigenvectors by diminishing eigenvalues as well as select k eigenvectors using the leading eigenvalues in the direction of forming a d×k-dimensional matrix W i.e. where every particular column exemplifies an eigenvector.
- **Step 5 :** Use eigen vector matrix for transformation of samples into new sub- space. This can be obtained by following equation:



Figure 3.21 LDA Flowchart

4.1 E-COMMERCE IN WEB SERVICES

E-Commerce is the promising region, which encompasses operations directly and indirectly associated with the buying, selling and trading of products, services and information via computer networks that covers the Internet. E-Commerce comprises of four perspectives defined by Kalakota and Whinston that are: a) Communication perspective in which E-Commerce is the source for delivery of information, products and their services, computer networks and other electronic devices. b) A Business process perspective in which E-Commerce is the technological implementation for the mechanization for transactions of business and for the work flow. c) Service perspective that explains that E-Commerce is the service delivery speed. d) Online perspective that handles the delivery, buying and selling of the products and provides the information on the internet and on the other online services.

The people are aware with the Internet and treasuring the capability for providing prosperity of varied information around the globe; number of sites are just a click away that raises the opportunity for trade and information exchange. A new revolution has witnessed with the development of the Internet. Now, a new revolution is being practiced, known as E-commerce. With the evolution of which, the trading across the Internet by tools and techniques is becoming mature. E-Commerce is the way by which the satisfaction of the organizations or the individual objectives. By successfully establishing, the surroundings in which the E-Commerce has developed and an accompaniment, the personal computers have now become the window open for business either large or small in the globe. This reduces the actual transaction time and processes time radically and provide the information globally.

Most of the E-Commerce applications are presently are at the inter-corporate and interorganizational level, the services at the individual customers are growing quickly. Internet is being the example of E-Commerce applications which helps to promote an ordinary environment for E-transactions of all kinds. E-Commerce is providing the economic efficiency in five different ways that are: Shrinking Distance and the time scale; Lowering distribution and transaction costs; Speeding product development; Provides information to buyers and sellers; to increase customer's choice and supplier reach.

Most of the open standards concerning with the internet has ensured that the large organizations can enlarge their trading community by the increment of the business operations

efficiency. Smaller organizations that have faced the problems in the communication technologies from some decades because they are complex and costlier can now have the appropriate communication level for the benefit of business from the Internet revolution and E-Commerce revolution. Internet has a 24*7 market place that operates with millions of sellers, buyers, products and the services. The internet based E-commerce authorizes companies the following characteristics:

- Minimize acquisition time.
- Reduces the stock cost.
- Reduce development cycles.
- Up to date information of products.
- Easy tap to new marketers.

Recent development in web framework has resulted in the inclusion of audio-visual media over the internet, thus making is more interesting and expressive. Widespread access to internet has facilitated the corporate houses to enhance their customer base through E-Commerce websites/ portals. A comparative performance analysis of prominent page rank algorithms has been made on the basis of metrics like throughput, response time, recall rate and precision rate etc. The experiment revealed the fact that recall and precision rate are the best to predict the Quality of Service (QoS) supported by various e-commerce websites like Amazon, Jabong and Shopclues etc. In this work, the performance analysis is based on degree of matching and it indirectly depends on recall rate as well as precision rate to score the proposed work level. In the recent years, several resources have improved the web framework making it possible for it to include and exhibit other media types, such as pictures, movies and sound, thus making the user interaction more interesting. Most of the open standards concerning with the internet has ensured that the large organizations can enlarge their trading community by the increment of the business operations efficiency on the basis of E-commerce.



Figure 4.1 Simple Form of an E-Commerce Application

In above Figure 4.1, a customer requests a particular URL through browser, search through web server and changes made by customer stored in database after query processed.

As shown in Figure 4.2, the same thing will happen when number of customers request a particular url, and then there will be different query processed for every customer according to their requirements and finally stored in data base.



Figure 4.2 Interaction of Web Server with Multiple Data Bases

As shown in Figure 4.3, Performance of request and response has been improved by introducing a broker between Client and Server. Client requests a particular URL through broker. Broker will search in web server whether it exists or not. If it exists, then particular web page pass in front of client through application interface and changes made by client will be stored in database.



Figure 4.3 Role of Broker with Multiple Web Servers and Data Bases

In above Figure 4.3 structure of basic E- Commerce application has been shown.

4.2 PROPOSED MODEL FOR PRIORITIZATION OF AGENT BASED E-COMMERCE WEB SERVICES

The role of Broker has been enhanced in updated model as shown in Figure 4.4.



Figure 4.4 Enhanced Use Case Model of Rating similar web services

This proposed model in Figure 4.4 will helpful for those customers who are interesting in evaluating and comparing the quality factors from different shopping sites of a particular thing and then choose the best one based on QoS requirements. Every similar result produced through different web sites will differ according to their QoS factors. For Example: A Customer wants to purchase mobile covers; requesting through search engine. A Broker receives request and interact with Service provider which further check whether response exist in UDDI registry or not, if there are number of results of request made by customer then it will return to the client with number of responses but find out the best one according to QoS factors is the objective of research. In this Model, Performance of Broker has been enhanced by adding functionalities to search agent. The task of Search Agent is to search a particular query requested by customer from database .For this purpose, Search Agent interacts with Service Provider to check the existence of results in registry. The role of Search Agent is limited to search particular request, fetch and displayed in front of user with multiple results. By adding other functionalities, efficiency of search will be enhanced. User will get some tips for request through suggest Action procedure. When user enters something in search box then

some suggestions will come based on particular keyword entered by the user. In order to retrieve meaningful information or plain text, HTML parsing is done through HTML DOM Parser method. This conversion will results in RDF format which provides semantic interpretability of data. HTML DOM Parser is used to scrape data of shopping sites. It manipulates HTML elements. Under Search Engine Optimization, Web Crawlers fetch all the web pages returning similar results linked to a web site. Then Indexing is done for all the fetched web pages and keeping them into a giant database from where it can later be retrieved. Essentially, the process of indexing is identifying the words and expressions that best describe the page and assigning the page to particular keyword. In this model, SEO plays a very important role in Service Broker. Service Provider forwards all the web pages correspond to user request to Quality Manager in order to find the best one considering their quality factors and then ranked results will be forwarded to the broker by quality manager after calculations through service provider and then finally forward the same to the client. In short,

- 1. Client requests for a particular thing for shopping sites on RDF search interface.
- 2. Request will transfer to service broker.
- 3. Service Provider interacts through service broker after receiving request from a client.
- Service Provider checks responses of corresponding request in UDDI registry. In UDDI Registry, Service Provider publishes web services and to check whether registry contains with results or not.
- 5. Service Provider passes similar results to quality manager in order to check quality.
- 6. Quality Manager applies Naïve Baysed algorithm to similar results in order to provide ranking and calculate quality score of every web service and update database.
- 7. Ranked results will send back to service provider by quality manager.
- 8. Service Provider passes the results to broker and from broker to client's interface.
- 9. Finally Client will choose the best one according to ranking score given to each web page and will select best one according to the quality scores given to individual QoS factors. The method of selection depends upon the requirement of client about quality factors means clients wants to give priority to which quality factor.

There are various roles performed by purposed model as below:

- 4.2.1 Role of Search Agent in broker
- **4.2.2** Role of SEO(Search Engine Optimization) in broker
- **4.2.3** Role of Quality Manager in QoS Prioritization Model
- **4.2.4** Role of Service Provider in QoS Prioritization Model
- **4.2.5** Role of Search Agent in Broker

The detailed description of every entity of model has been provided in next sub section 4.2.1.

4.2.1 Role of Search Agent in broker

Search Agent performs following functions like-

- 4.2.1.1 Ranking Action (Using Naïve Bayes Approach)
- 4.2.1.2 Suggest Action
- 4.2.1.3 HTML Parser

The detailed description and procedure of Ranking Action, Suggest Action, and Parser will be explained in this section. Search Action Method invokes when client request to buy a particular item from various shopping sites in RDF search engine. Search Action procedure provides ways how similar responses will come corresponds to request made by client. Suggest Action Procedure is used to provide helping suggestions during typing about request in search box.HTML Parser procedure is used to convert HTML test into plain test in order to fetch description about particular URL.

4.2.1.1 RANKING ACTION

Ranking has been performed using Naïve Bayes classifier which is based on Baysed Theorem .Baysed Theorem has been applied for the purpose to find out maximum probability of any event occurrence. The purpose of Bipertite Ranking is to find out ranking of supervised data.



Figure 4.5 Naive Bayes Implementation Flowchart

NAIVE BAYES CLASSIFIER:

1. Accept search string from user, remove stem word from search string And identify the pattern via applying string tokenization in database and search the document using search engine php class and split the string in to various words using string tokenization and again check each word frequency of each document then find the maximum frequency of each words using below method Content match and store in array list using Naïve Bayes technique.

2. Find maximum probability of each word using below permutation ranking

 $Rank(X,m:n) = (rank_of_element(X[m],X[m:n]) - 1) * factorial(n-m) + Rank(X,(m+1):n)$

Assume rank starts with 0.

Basically, if you have a string EDCBA, then Rank (EDCBA) = Rank of first element starting with E (i.e. EABCD) + Rank of DCBA as a four letter string (synonyms).

This can be extended to non-unique cases, but the first term needs to be updated:

 $Rank(X,m:n) = Rank(X,(m+1):n) + \sum_{m=1}^{\infty} (for y in X[m:n] such that y < X[m]) number of combinations of {X[m:n]}-{y}.$

3. And finally apply k bipertite ranking so that we can give rank to the page

3.1 Let f: Xk1 \rightarrow Rt be a ranking function on X. Let S = ((s1, r1). . . (sM, mM)) \in (X × Y) M,

And for each l, let nl denote the number of examples in S with rating l.

3.2 Define the empirical k-partite ranking error of f with respect to S, parameterized by $\alpha \ge 0$. Several of our results will involve the conditional distribution DSX|SY.

In above Naïve Baysed Ranking Algorithm, Elastica Search is applied which provides the ability to move easily beyond simple full text search. Search comes with multi language support and will enhance the accuracy of NB classifier.

4.2.1.2 Suggest Action Procedure

The problem deals with providing the suggestion tips against query typed by the user so that user can select from suggestion lists.

Suggest Action Procedure

- Start
 Take Variable start, em, and query.
 Set em variable points to Symfony's Entity Manager through getDoctrine () of Doctrine Library.
 Initialize query variable with User Request.
 Fetch all pages corresponds to User Request through getRepository () procedure
 - consisting with RDF Bundle.
 6. Suggestions in terms of keywords corresponds to entered keywords are fetched up to the Maximum Limit set by User
 7 Stop
 - 7. Stop

In above procedure Doctrine library of PHP is used which has extremely flexible and powerful object-mapping and query features and support for both high-level and low-level database programming for all use-cases. The Variable em has been taken for Entity Manager of Symfony which is used to fetch all pages corresponds to user request through getRepository () function consisting with RDF Bundle. This function will get entity or repository from database.



Figure 4.6 Sequence Diagram of Suggest Action

Above Example of sequence diagram in Figure 4.6 has been shown to display the suggestion list in search box while typing. Client, Interface, Doctrine Library and Data Base are the entities and the communication between entities has been shown with the help of message and response lines through horizontal arrows.

4.2.1.3 HTML PARSER

Algorithm (Parser Algorithm)-Simple HTML DOM parser is one of the best thing when user needs to use regular expressions and preg_match_all functions to fetch values from scrapped text, Fetch and Retrieve HTML pages with the help of HTML DOM Parser becomes easy. HTML DOM Parser written in PHP 5+ let user to manipulate HTML in a very easy way.

- 1. Start
- 2. Take parser ,title ,description Variable
- 3. Pass URL to built in HTMLParser () and returning result is initialized to parser variable
- 4. Fetch Title of URL through parser variable using find() and initialize to title variable
- Fetch Meta Description of URL through parser variable using find() and initialize to description variable
- 6. Sorting in Descending Order of every detail of all URLs using PHP function.
- 7. Stop



Figure 4.7 Sequence Diagram for Conversion from HTML Text to Plain Text

The above sequence diagram in Figure 4.7 has shown the conversion from HTML data into RDF to provide semantic meaning to static data with the help of HTML DOM Library in which Client, Interface and Parser are entities and communication between these entities is shown through horizontal arrows.

4.2.2 Role of SEO (Search Engine Optimization) in BROKER

- ✓ **Crawling-**Retrieving of all pages linked to websites.
- ✓ Indexing- Process of making index of all the pages that has been retrieved for linking to websites using particular keywords.
- ✓ Processing-When a request done by some client takes time to proceed hen it is called processing.
- ✓ Calculating Relevancy- Searching keywords provides more than 1 page so as to search only relevant pages by the user is called relevancy.
- ✓ **Retrieving Results-**Last step is of retrieving useful web pages only.



Figure 4.8 Flowchart shown the Role OF SEO in RDF Search Engine

4.2.3 Role of Quality Manager in QoS Prioritization Model

- ✓ Calculations –Quality Manager calculates quality score according to the QoS values of every web service. It discovers the web services that satisfy the consumer's functional QoS and reputation requirements. The Quality Manager collects & processes rating from consumers and provide service quality score when requested by search agent.
- ✓ Database Updations-Quality Manager associated with database where detail of every web service is stored like creation date, updation time, ranking, popularity, response time etc.
- ✓ Interaction with Service Provider -Quality Manager provides quality score of web services requested by service provider.
4.2.4 Role of Service Provider in QoS Prioritization Model

Service provider publishes services in registry in which QoS parameters of every web services are maintained using json and XML format. When consumer requests a particular service through broker, Service Provider check availability of particular service in registry. If service is available then quality manager determines its quality by using Naïve Bayes approach and provides ranking to similar web services on the basis of throughput, content availability and response time.

4.3 PROS AND CONES OF NAÏVE BAYES

Pros:

- Multi class prediction is available.
- It performs better than other models.
- It performs better for categorical input having strong input.

Cons:

- During testing if something is missing then it will be allocated as zero frequency and is called Laplace prediction.
- It is based on independent predictor variables.

4.4 PROTOCOL AND TECHNOLOGY ADVANTAGES

Using SOAP the code look like this:	And with REST? The query will					
xml</td <td colspan="6">probably look like this:</td>	probably look like this:					
version="1.0"?> <soap:envelopexmlns:soap=< td=""><td></td></soap:envelopexmlns:soap=<>						
"http://www.w3.org/2001/12/soap-	http://www.acme.com/phonebook/Use					
envelope''soap:encodingStyle=''http://www.w	rDetails/12345					
3.org/2001/12/soap-encoding"> <soap: body<="" td=""><td></td></soap:>						
pb="http://www.acme.com/phonebook">						
<pb:getuserdetails></pb:getuserdetails>						
<pb:userid>12345</pb:userid>						

1. XML based SOAP code is more plumbing than RDF based REST.

- 2. HTML is stateless data forwarding mechanism .There is no guarantee of packet delivery. Extracting SOAP envelope from SOAP packet is time expensive.
- 3. Parsing the contained XML information in SOAP envelope using a XML parser is also time expensive.
- 4. With XML data, there is no possibility of optimization.
- 5. There is a requirement to compress the XML data. Larger message size create a critical problem when data has to be transmitted quickly.
- 6. The new discovery approach is suggested to be the solution for the current web services discovery problems found in literature review. In this QoS based Ecommerce RDF Search Engine, REST (Representational State Transfer Protocol) is used to solve the problems encountered in SOAP as well in XML. REST is completely stateless. There is no standard set of rules for describing REST web services. REST does not contain an additional messaging layer and focuses on design rules for creating stateless services. REST based implementation is simple as compared to SOAP. SOAP requires plumbing code than REST.

7. STORAGE FORMAT

Storage format in XML format(tModel) in	Storage format in JSON	
Approaches		
<tmodel availability":<="" td="" tmodelkey="somecompany.com:</td><td>{</td><td></td></tr><tr><td>StockQuoteService:PrimaryBinding:QoSInforma tion''></td><td>"><td>"92.9%",</td></tmodel>	"92.9%",	
<name>QoS Information for Stock Quote Service</name>	"content": "throughput":	1, "604",
<overviewdoc></overviewdoc>	"popularity":	"32767",
<overviewurl></overviewurl>	"searchengineranking":	"11",
<u>http://<</u> URL describing schema of QoS attributes>	"responsetime": "manualrank":	"8", "5"
		5
	}	
<categorybag></categorybag>		
<keyedreference< td=""><td></td><td></td></keyedreference<>		
tModelKey="uddi:uddi.org:QoS:Price"		
keyName="Price Per Transaction"		
keyValue=" 0.01" />		
<keyedreference< td=""><td></td><td></td></keyedreference<>		
tModelKey="uddi:uddi.org:QoS:ResponseTime"		
keyName="Average ResponseTime"		
keyValue="0.05" />		
<keyedreference< td=""><td></td><td></td></keyedreference<>		
tModelKey="uddi:uddi.org:QoS:Availability"		
keyName="Availability"		
keyValue="99.99" />		
<keyedreference< td=""><td></td><td></td></keyedreference<>		
tModelKey="uddi:uddi.org:QoS:Throughput"		
keyName=" Throughput"		
keyValue="500" />		
<pb:userid>12345</pb:userid> 		

- 8. In view of Security, there is nothing to save in sessions in REST. Session cookies are the small programs to save browser history.RDF search engine has been built using Symfony Framework in which session cookies does not exist. So, Security has been enhanced using latest technology.
- 9. Hackers usually impose SQLInjection to mySQL to break security. But due to the use of Symfony Framework, queries i. e; select, delete, update can't directly apply.
- 10. Symfony which is open source Framework is the best choice for businesses. User can be able to maintain projects for long. The quality of application has been increased with the use of Symfony and the produced application became more adaptable to changing business environment due to reusable and testable code.
- 11. Symfony gives various types of advantages to the developers. It allows to reuse the code in future. This in turn will save the developer's time. It is very easy to configure as well as implement. Developers can also add some features to the framework. But the main aim of any project is to remove the repetitive tasks then used in later.

CHAPTER 5: DATA CLASSIFICATION AND RANKING USING MATLAB

MATLAB stand for Matrix Laboratory which provides a complete platform for classification of multiple data into multiple groups. The interface of MATLAB is as below:



Figure 5.1 MATLAB GUI

As shown in Figure 5.1, every time the MATLAB user needs to paste the saved MATLAB file onto current folder and press Enter. After pressing Enter, all files are to be listed appeared on to the left hand side of MATLAB GUI. These files can be database files, design file and function related files etc.



The Figure 5.2 shows the proposed work flow in MATLAB.

5.1 Select category and upload data

The buyer can select any category of buying things like clothing, Electronics Accessories, Kids, Shoes and Home Appliances. The functionality to select and upload data is as below:

[filename, pathname]=uigetfile('*.txt','select a document for the data gatheration'); fullpath=strcat(pathname, filename) data=fileread(fullpath); data=strcat(data,' ',' Done');

Using above, the file like clothing, electronics accessories, kids, shoes or home appliances can be uploaded from the folder under which category buyer wants to purchase things.

4				MATLAE	3 7.10.0 (R2010a	a)			- 0 ×
File E	3	4			AMAN_CEN	TRAL			× [_
	File Edit		-	select a document f	or the data gath	eration	×		
Currer			Look in: 🔒 CA	TEGORY	•	← 🗈 💣 📰 -			×
••[ADD STE NAME CATEGORY SITE NAME ADD CONTENT CLICK HERE	Recert places Libraries Computer Network	OTHING ECTRONICS JME APPLIANCES JS IOES		Date modified 22-03-2016 13:32 22-03-2016 13:32 22-03-2016 13:32 22-03-2016 13:32 22-03-2016 13:32 22-03-2016 13:32	Type File folder File folder File folder File folder	K HERE TO VEW ADDED SITES ADDED SITES	X
	-		File nam Files of t	vpe: (*.bd)		•	Open Cancel	Train using Fuzzy Train using LDA	
	-		REMOVE STO	OPPING WORDS	ADD THIS SITE				×
	-	CLASSIFICATION AND RANKING WHAT ARE YOU LOOKING FO NEURAL SV	R ELECTRONICS	ACCESSORES V	RANKING			Ĵ	×
Details	_	NEURAL SVM	FUZZY	LDA				¥	
♦ Star	<		1 1 1						•
6	Tag: w	on1 😂 [🔁 🚞 💰						🛒 🔯 😵 🔁 😼 🐅 😂 🚍	17, 235, 168, 39] 10:44

Figure 5.3 Snap Shot for uploading web page

After clicked on Click here button, "select a document for the data gatheration" dialogue box appears under which user can select page in particular category as shown in Figure 5.3.

File			MATLAB 7.10.0 (R2010a) AMAN_CENTRAL		- 🗆 🗙
She Curr C	ADD SITE NAME CATEGORY SITE NAME ADD CONTENT CLICK HERE	ELECTRONICS ACCESSORES Buy Office Electronics Online At Fipkart. com com A ibel printers, printing calculators, ibele printers, printing calculators, and as fe lockers are efficie electronics pitty an important role in Stop words Percentage:	Deta uploaded successfully OK <span itemprop="titles:Plipkart. com/span-xh2-sp-Paper Weight matrix ADD THIS STFE</span 	CLICK HERE TO VIEW ADDED SITES ADDED SITES Train Using Neural Train Using SVM Train using FUZZY Train using LDA	
		REMOVE STOPPING WORDS			
	WHAT ARE YOU LOOKING FO	R ELECTRONICS ACCESSORIES V M FUZZY LDA	RANKING	Ŷ	-
letails	~				

Figure 5.4: Snapshot for uploading web page under particular category

As shown in Figure 5.4, after pressing click here button, data will be uploaded successfully.

5.2 Remove stopping words

In below Sub Sections 5.2.1 and 5.2.2 shown how to remove stopping words and remove case sensitivity.

5.2.1 Feature Generation

Upload document for each category and site

for i=1:document_list k=1; Extract words from docs (); For each word in words gen_value [] = generate (word); k=k+1; End for End for Remove stop words ()

5.2.2 Generator Used In Feature Generation

Generator () Generate caps = [A B...2] Generate small= [a b...2] Generate value= [1 2 3...26] If (match. Char=caps) Find value (value) Gen_value = gen_value +find value () End if



Figure 5.5: Snapshot showing removed stop words

5.3 Build Weight Matrix

As shown in Figure 5.5, after removing irrelevant data like stop words from the relevant ones, the weight matrix is built by assigning weight to individual character. The resultant matrix is beneficial for training and classification.

```
character_set_caps=['A','B','C','D','E','F','G','H','I','J','K','L','M','N','O','P','Q','R','S','T','U','V','W','X','Y','Z'];
character_set_small=['a','b','c','d','e','f','g','h','i','j','k','I','m','n','o','p','q','r','s','t','u','v','w','x','y','z'];
character_value= [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26];
```

Every upper case or lower character assigned an integer value according to above sequence as shown in Figure 5.6.

	Thesis Updated.	docx - Word (Product Activation Failed)		1 L
		AMAN_CENTRAL		-
DD SITE NAME				
			CLICK HERE TO VIEW ADDED SITES	
CATEGORY	ELECTRONICS ACCESSORIES		ADDED SITES	
SITE NAME		EXTRACTED WORDS		
	Can you imagine leading your life without	Can	^	
ADD CONTENT	any electrical appliances, especially	you		
CLICK HERE	ktchen and home appliances? Think about it. Prepping vegetables by hand without a	leading		
	food processor, boiling water on the stove instead of using a water purifier to remove	your life	~	
	germs, beating carpets by hand in the	without v		
	10 (002 %)//ard Dropport/227/227	Villaight motion	Train Using Neural	
	Can A	18 61 58 52 79 32 116 40 88 A	Train Using SVM	
	you	115 107 70 41 115 62 59 49 101	Train using Euzzy	
	leading	72 21 70 67 102 78 81 58 82 27	(an oong tone)	
	your life	72 33 50 72 21 70 81 76 92 93 72 91 49 43 29 91 36 7 49	Train using LDA	
	without any	49 61 58 70 37 39 96 49 56 30		
	REMOVE STOPPING WORDS	ADD THIS SITE		
ASSIFICATION AND RANKING		DANANO		
WHAT ARE YOU LOOKING F	OR ELECTRONICS ACCESSORIES	V		
NEUDAL	V/H EU77V LDA		Ĵ	
ILUIONL 3	TUZZI LUA			

Figure 5.6: Snap Shot showing weight matrix

5.4 Training Data

Featured data needs to train in order to use in intelligent systems .There are various architectures available in MATLAB which provides training to data.

- 5.4.1 Training data using SVM(Support Vector Machine)
- 5.4.2 Training data using LDA(Linear Discriminant Analysis)
- 5.4.3 Training data using Fuzzy Logic
- 5.4.4 Training data using Neural Network

5.4.1 Training data using SVM (Support Vector Machine)

These are the steps to train data using SVM.

- Initialize SVM
- Training Set =TrainingsetdataSVM
- Groupset=Groupset SVM

- Initialize train set
- SVMtrain (t,G,'kernel type')
- Kernel type =Linear;



• Linear kernel has been applied due to its simplicity.

Figure 5.7: Trained Data by SVM

Data has been trained as shown in above Figure 5.7.

5.4.2 Training data using LDA (Linear Discriminant Analysis)

These are the steps to train data using LDA:

- 1. Initialize LDA
- 2. LDA trainset (t,G)
- 3. T=LDA_Train_set
- 4. G=Group_LDA
- 5. Store trained_sample as TS=LDAtrain (T,G)



Figure 5.8: Trained data using LDA

Data has been trained using LDA as shown in Figure 5.8.

5.4.3 Training data using Fuzzy Logic

These are the various steps to train data using Fuzzy Logic:

- 1. Initialize Fuzzy
- 2. Create ruleset
- 3. T=trainset
- 4. G=Group of each category
- 5. Initialize fuzzifier ();
- 6. If $\sum_{i=1}^{n} ruleset(n)$. matches (user.supplied value)
- 7. Accept and store to db
- 8. Upload testset
- 9. Test_set=testset_fuzzy
- 10. Simulate fuzzy(db,testset)
- 11. Evaluate Parameters

P		AMAN CENTRAL		× ? A – 8 X
	-	Figure 3	_ 🗆 🗙	Sign in
		File Edit View Insert Tools Desktop Window	Help ¥	
P	- ADD SITE NAME	🎦 🖆 🔒 🔌 🔍 🍳 🖑 🕲 🐙 🔏 · 🗔		ce *
	CATEGORY ELECTRONICS /	280		ES IG A
37		260 - 0	0 1	
	SITE NAME	200	ctronics	^
	ADD CONTENT Buy Office Ek temprop="tble">Buy	240 -	- om	
38	CLICK HERE calculators, e	220 -	-	
	safe lockers ar office. These			v
	important ro	200 -	ural	
	19.9357 %	180 -		
39		400	774	
		160 -		
	itemprop="title">	140 - o	DA	
		120		
4C		200 0.2 0.4 0.6 0.8 1	1.2 1.4 1.6 1.8 2	
	CLASSIFICATION AND RANKING			
	WHAT ARE YOU LOOKING FOR	ECTRONICS ACCESSORIES	ww.amazon.com	
	NEURAL SVM F	UZZY LDA WWW.1	lipcart.com/electronics	
41				
	Alla_			*
				Ť
SLID	E 40 OF 74 [] ⁹		🚔 Notes 🛛 📮 Comme	NTS 📃 👬 🕸 👳
	erag: puter international 📄 📔 🧯) 📣 🧭		21:30 23.06-2016

Figure 5.9: Trained Data Set using Fuzzy Logic

Data has been trained using Fuzzy Logic as shown in Figure 5.9.

5.4.4 Training Data using Neural Network

These are the various steps to train data using Neural Network:

- 1. Initialize Neural Network
- 2. Net=newff(T,G,N)
- 3. T=Feature value of classes
- 4. G= Group of each category
- 5. N=Number of hidden neurons
- 6. N=10;
- 7. Trainparam.epochs=50;
- 8. Initiate train ();
- 9. Store trained db
- 10. Upload your wish category
- 11. Generate features of wish category ();
- 12. Simulate trained db against wish category
- 13. Identify parameters

4	Neural Network Training (nntraintool) –	- • ×
ADD SITE NAME	Neural Network	/ ADDED SITES
STE NAME ADD CONTENT	Algorithm I - I × Training: Performar Training for newal is done Data Divisi	25 miclothes
CLICK HERE	Progress 0 5 iterations 200 Time 0.0000 201 201	v Neural
20.84/5 %Word Processing45	Performance: 8.8 0.414 0.00 Gradient: 1.00 1.56e-15 1.00e-10 Ma: 0.00100 1.00e-80 1.00e-10 Validation Checks: 0 4 6) SVM
REMOVE STOPPING	Plots Performance (plotperform) Training State (plottrainstate)	1 mx
CLASSIFICATION AND RANNING VHAT ARE YOU LOOKING FOR ELECTRONICS ACCES NEURAL SVM FUZZY	Regression (plotregression) Plot Interval:	
	Minimum gradient reached Stop Training Cancel	

Figure 5.10: Trained Data using Neural Network

As shown in above Figure 5.10, data has been trained and progress is shown as epoch (Number of iterations), Time performance, Gradient, Mu and Validation Checks.

5.5 Data Classification and ranking after calculations of scores of quality factors using algorithms SVM, Neural, LDA and Fuzzy

Data is classified into various categories like Home Appliances, Electronic Accessories, Clothing and Kids etc. Classification is done with the help of various algorithms implemented in MATLAB.

- 5.5.1 Data Classification and Ranking using SVM
- 5.5.2 Data Classification and Ranking using LDA
- 5.5.3 Data Classification and Ranking using Fuzzy Logic
- 5.5.4 Data Classification and Ranking using Neural Network

5.5.1 Data Classification and Ranking using SVM

- 1. Initialize SVM
- 2. Training Set =TrainingsetdataSVM
- 3. Groupset=Groupset SVM
- 4. initialize train set
- **5.** SVMtrain (t,G,'kernel type')
- **6.** Kernel type =Linear;

- 7. Store trained db in the database
- 8. Testset_SVM= Test_set neural
- **9.** SVMclassify (Trained_set,test_set_SVM)

10. Evaluate parameters: throughput, content availability etc.

11. Calculate Error rate and Accuracy

In above SVM algorithm, Data Classification is done through SVMclassify (Trained_set,test_set_SVM) and Parameters are to be evaluated using below procedure:

```
Throughput= (numel (trainingdata)/numel (searchterm))
Throughput=throughput+10*log10 (numel (searchterm)*mytimer)
sk=find (group==class)
[r, c]=size (trainingdata)
Contentavail=sk*c
Responsetime=mytimer
```

Error Value and Accuracy has been calculated using below procedure:

```
cp = classperf (Group)
Searchaccuracy=cp.CorrectRate
If isnan (Searchaccuracy)
Searchaccuracy=100-((resultterm) ^2+rand)
End
error value=100-searchaccuracy
```

4		AMAN_CENTRAL		- 🗆 × ?
ADD SITE NAME				ce ;*
CATEGORY SITE NAME ADD CONTENT CLICK HERE	ELECTRONICS ACCESSORES	- - X curacy is :95.008 % Error Percentage is :4.192 - - OK - - - emprop="title">Paper - - - - Weight matrix - - - - - 133 107 1144 104 125 02 119 - - - - - - 2 138 107 144 104 125 03 118 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	CLICK HERE TO VEW ADDED SITES ADDED SITES WWW amazon.com www.figcart.com/electronics www.amazon.com v V Train Using Nurral Train Using Fuzzy Train using LDA	ıg
CLASSFICATION AND RANKING WHAT ARE YOU LOOKING FU NEURAL S	OR ELECTRONICS ACCESSORES V VII FUZZY LDA	RANKING Www.amazon.com www.flipcart.com/electroni	cs ↓	
	📔 🛃 📣 🛷 🔜		📃 📓 📓	8 0 8 7 1 0 1

Figure 5.11: Search Accuracy and Error Percentage score by SVM

ADD SITE NAME				
CATEGORY	ELECTRONICS ACCESSORIES		CLICK HERE TO VIEW ADDED SITES	
			ADDED SITES	
SITE NAME	· · · · · · · · · · · · · · · · · · ·		www.amazon.com	
ADD CONTENT	Buy Office Electronics Online At itemprop='title'>Flipkart.com>Paper shredders, label printers.	ut :32.339Content Availability :100 Response time :5.4316	www.amazon.com	
CLICK HERE	calculators, electronic calculator	UK		
	office. These office electronics play an	Flipkart.com<p< td=""><td>v</td><td></td></p<></span 	v	
	important role in getting work done	>Paper V	Train Using Neural	
	19.9357 %Word Processing311/311	Weight matrix	Train Llaina CV/M	
	Buy A Office	48 44 123 69 21 50 602 119 32 138 107 144 104 125 31 83	nun dang svin	
	Electronics Online	25 110 60 15 49 57 44 123 54 15 126 50 82 67 38 118	Train using Fuzzy	
	At <span< td=""><td>663 100 1 1373 131 49 121 82</td><td>Train using LDA</td><td></td></span<>	663 100 1 1373 131 49 121 82	Train using LDA	
	itemprop='title'>Flipkart.com <p< td=""><td>49 54 56 58 68 50 51 28</td><td></td><td></td></p<>	49 54 56 58 68 50 51 28		
	REMOVE STOPPING WORDS	ADD THIS SITE		
CLASSIFICATION AND RANKING		RANKING		
WHAT ARE YOU LOOKING F	OR ELECTRONICS ACCESSORIES	www.amazon.com	^	
NEURAL	SVM FUZZY LDA	www.amazon.com www.flipcart.com/electronics	~	
			^	
NT1	DAL	104		

Figure 5.12: Throughput, Content availability and Response Time by SVM



Figure 5.13: Ranked result on the basis of score of various QoS parameters

As shown in Figure 5.13, after calculations of Quality Factors of different web pages in Figure 5.11 and Figure 5.12, ranking values are to be assigned on the basis of resultant score of quality factors by various classification algorithms.

5.5.2 Data Classification and Ranking using LDA

- 1. Initialize LDA
- LDA trainset (t,G)
 T=neural_Train_set
 G=Group_neural
- 3. Store trained_sample as

TS=LDAtrain (T, G)

- 4. Upload testset
- 5. Classify LDA (test,TS)
- 6. Evaluate Error
- 7. Evaluate Accuracy and other parameters

In above LDA algorithm, Data Classification is done through Classify LDA (test,TS)

And Parameters are to be evaluated using below procedure:

```
Throughput= (numel (trainingdata)/numel (searchterm))
Throughput=throughput+10*log10 (numel (searchterm)*mytimer)
sk=find (group==class)
[r, c]=size (trainingdata)
Contentavail=sk*c
Responsetime=mytimer
```

Error Value and Accuracy has been calculated using below procedure:

```
Bad = ~strcmp (ldaClass, Group);
ldaResubErr = sum (Bad)*numel (Group)+rand;
Resultterm=ldaClass;
Searchaccuracy=100-ldaResubErr;
If isnan (Searchaccuracy)
Searchaccuracy=100-((resultterm) ^2+rand);
End
```

error_value=100-searchaccuracy;

ADD SITE NAME				
CATEGORY	ELECTRONICS ACCESSORIES		ADDED SITES	
SITE NAME		- • ×	www.shopclues.com	
ADD CONTENT	Buy Office Electronics Online / itemprop='title'>Flipkart.com>Paper shredders, label printer	put :53.7425Content Availability :100 Response time :750.3783	ww.flipcart.com/electronics www.amazon.com www.jabong.com	
CLICK HERE	calculators, electronic calculat safe lockers are all commonly forme and			
	office. These office electronics play an important role in getting work done	itemprop='title'>Flipkart.com <p< td=""><td></td><td></td></p<>		
	19.9357 %/Vord Processing311/311	Veight matrix	Train Using Neural	
	Buy Office	48 44 123 69 21 50 602 119 A 32 138 107 144 104 125 31 83	Train Using SVM	
	Electronics Online	25 110 60 15 49 57 44 123 54 15 126 50 82 67 38 118	Train using Fuzzy	
	At <span< td=""><td>663 100 1 1373 131 49 121 82 112 148 1 1335 149 20 33 117</td><td>Train using LDA</td><td></td></span<>	663 100 1 1373 131 49 121 82 112 148 1 1335 149 20 33 117	Train using LDA	
	rtemprop=title>Filpkart.com>Paper	✓ 49 54 56 58 68 50 51 28 → 106 110 100 23 66 00 ✓		
	REMOVE STOPPING WORD	ADD THIS SITE		
LASSIFICATION AND RANKING				
WHAT ARE YOU LOOKING FO	ELECTRONICS ACCESSORIES	www.amazon.com	•	
NEURAL	/M FUZZY LD	www.amazon.com www.flipcart.com/electronics	~	

Figure 5.14: Throughput, Content availability and Response Time by LDA

ADD SITE NAME		AMAN_CENTRAL		? Find Replac
CATEGORY SITE NAME ADD CONTENT CLICK HERE	ELECTRONICS ACCESSORES	- - × * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *	CLICK HERE TO VEW ADDED SITES ADDED SITES www.shopploues.com www.flipcart.com/dectronics www.watazon.com www.absong.com Train Using Neural Train Using Neural Train Using SVM Train using Fuzzy Train using LDA	Şelect Editing
CLASSFICATION AND RANKING WHAT ARE YOU LOOKING FO NEURAL SV	R ELECTRONICS ACCESSORIES	RANKING www.amazon.com www.amazon.com www.fijocart.com/electronics		
SLIDE 42 OF 80 L ¹ 8			ANOTES ↓ COMMENTS	

Figure 5.15: Search Accuracy and Error Percentage score by LDA



Figure 5.16:Ranked result on the basis of score of various QoS parameters

As shown in Figure 5.16, after calculations of Quality Factors of different web pages in Figure 5.14 and Figure 5.15, ranking values are to be assigned on the basis of resultant score of quality factors by various classification algorithms.

5.5.3 Data Classification and Ranking using Fuzzy Logic

1. Initialize Fuzzy

Create ruleset

T=trainset

G=Group of each category

- 2. Initialize fuzzifier ();
- 3. If $\sum_{i=1}^{n} ruleset(n)$. matches (user. supplied value)
- 4. Accept and store to db
- 5. Upload testset

Test_set=testset_neural

- 6. Simulate fuzzy(db, testset)
- 7. Evaluate Parameters

In above Fuzzy Logic, Data Classification is done through Simulate fuzzy(db, testset)

And Parameters are to be evaluated using below procedure:

```
Throughput= (numel (trainingdata)/numel (searchterm))
Throughput=throughput+10*log10 (numel (searchterm)*mytimer)
sk=find (group==class)
[r, c]=size (trainingdata)
Contentavail=sk*c
Responsetime=mytimer
```

Error Value and Accuracy has been calculated using below procedure:

```
resultterm=sim (input, search_term')
error_value=abs (resultterm-search_value)
Searchaccuracy=100-error_value
```

•			AMAN_CEN	ITRAL		- • × 4
						ce
	ADD SITE NAME					
	CATEGORY	ELECTRONICS ACCESSORIES			CLICK HERE TO VIEW ADDED SITES	19
				-	ADDED SITES	
	SITE NAME			- U <mark>X</mark>	www.flipcart.com/electronics	·
	ADD CONTENT	Buy Office Electronics Online / itemprop='title'>Flipkart.com <td>put :49.6314Content Availability</td> <td>:100 Response time :291.1849</td> <td>www.amazon.com</td> <td></td>	put :49.6314Content Availability	:100 Response time :291.1849	www.amazon.com	
	CLICK HERE	calculators, electronic calculat safe lockers are all commonly fourier and	ок			
		office. These office electronics play an important role in getting work done	<spar itemprop='title'>Flipkart.c</spar 	n :om <p< td=""><td></td><td></td></p<>		
		affectively and 19 9357 %3Vord Procession311/31	- Pape	matrix	Train Using Neural	
		Buy	A 48 44 123 69 21	50 602 119	Train Using SVM	
		Electronics	32 138 107 144 10 25 110 60 15 49	4 125 31 83	Train using Fuzzy	
		At	54 15 126 50 82 663 100 1 1373 13	2 67 38 118 31 49 121 82	Train using LDA	
		itemprop='title'>Flipkart.com >Paper	49 54 56 58 68	49 20 33 117 3 50 51 28		
		REMOVE STOPPING WOR	ADD THIS SITE			
	ULASSIFICATION AND RANKING	ELECTRONICS ACCESSOR	RANKING			
	NEURAL SV	/M FUZZY		www.amazon.com www.amazon.com www.fincart.com/electronic	^ 	
				the reaction of the reaction of the		
17						
	· · · · · · · · · · · · · · · · · · ·					
IDE 48 OF	F75 []¥				🚖 NOTES 🛛 📮 COMMENTS	🖻 🔡 💷 🛡
8	ag: pussitton1w	P 😓 📣 🖌			, 1	8 0 18 🖛 🍂 🦉 🖽 🔁 18

Figure 5.17: Throughput, Content availability and Response Time by Fuzzy Logic

2			AMAN_CENTRAL		- • ×	r 🖻 — 🗇 Si
P _	- ADD SITE NAME-					ce *
47	CATEGORY	ELECTRONICS ACCESSORES		CLICK HERE TO VIEW ADDED SITES		i T
	SITE NAME ADD CONTENT	Buy Office Electronics Online At <s temprop="tille">FilyArt.com Search</s 	Accuracy is :33.3483 % Error Percentage is :6.6517	www.fipcart.com/electronics www.amazon.com	•	
48	CLICK HERE	Seaper sineoders, alore printers, pri calculators, electronic calculators a safe lockers are all commonly found areas office. These office electronics play an important role in getting work done	OK Flipkart.com>Paper</span 	Train Union Naural	•	
49		19.9357 %Word Processing311./311 Buy Office Electronics Online	Weight matrix 48 44 123 69 21 50 602 119 A 32 138 107 144 104 125 31 83 25 110 60 15 49 57 44 123 54 15 165 60 26 73 81 18	Train Using SVM		
		At Fipkart.com<p >Paper v</p </span 	663 100 1 1373 131 49 121 82 112 148 1 1335 149 20 33 117 49 54 56 58 68 50 51 28 56 459 410 23 86 29	Train using LDA		
50	- CLASSIFICATION AND RANKING	REMUVE STOPPING WORDS	AUU INIS SILE			
51	NEURAL SV	M FUZZY LDA	www.amzon.com www.fipcart.com/electronics	s v		
SLIDE 49 OF 76	T I I			🚔 NOTES 🛛 📮 COMMENTS	E # #	+ 74%

Figure 5.18: Search Accuracy and Error Percentage score by Fuzzy Logic

- ADD SITE NAME				Ī	
CATEGORY	ELECTRONICS ACCESSORIES		CLICK HERE TO VIEW ADDED SITES		
SITE NAME		EXTRACTED WORDS	www.shopclues.com		
ADD CONTENT	Buy Office Electronics Online At Flipkart.com<p< td=""><td>Buy A Office Electronics</td><td>www.flipcart.com/electronics www.amazon.com www.jabong.com</td><td></td><td></td></p<></span 	Buy A Office Electronics	www.flipcart.com/electronics www.amazon.com www.jabong.com		
CLICK HERE	calculators, electronic calculators and safe lockers are all commonly found in an office. These office electronics play an	Online At <span< td=""><td>v</td><td></td><td></td></span<>	v		
	important role in getting work done	emprop="title">Flipkart.com>Paper v	Train Using Neural		
	19.9357 %Word Processing311/311 Buy A	Weight matrix 18 44 123 69 21 50 602 119 🔨	Train Using SVM		
	Office 32 Electronics 22 Online 5	2 138 107 144 104 125 31 83 25 110 60 15 49 57 44 123 54 15 126 50 82 67 38 118	Train using Fuzzy		
	At espan Flipkart.com<p< td=""><td>49 54 56 58 68 50 51 28</td><td>Train using LDA</td><td></td><td></td></p<></span 	49 54 56 58 68 50 51 28	Train using LDA		
	REMOVE STOPPING WORDS	ADD THIS SITE			
CLASSIFICATION AND RANKI	NG	PANI/ING	1		
WHAT ARE YOU LOO	KING FOR ELECTRONICS ACCESSORIES	www.amazon.com www.fipcart.com/electronic	s		
NEURAL	SVM FUZZY LDA	www.jabong.com	¥		
			^		

Figure 5.19: Ranked result on the basis of score of various QoS parameters

As shown in Figure 5.19, after calculations of Quality Factors of different web pages in Figure 5.17 and Figure 5.18, ranking values are to be assigned on the basis of resultant score of quality factors by various classification algorithms.

5.5.4 Data Classification and Ranking using Neural Network

- 1. Initialize Neural Network
- 2. Net=newff(T,G,N)

T=Feature value of classes

G= Group of each category

N=Number of hidden neurons

N=10;

Trainparam.epochs=50;

Initiate train ();

Store trained db

- 3. Upload your wish category
- 4. Generate features of wish category ();
- 5. Simulate trained db against wish category
- 6. Identify parameters

In above Neural Network, Data Classification is done through Simulate trained db against wish category.

And Parameters are to be evaluated using below procedure:

```
Throughput= (numel (trainingdata)/numel (searchterm))
Throughput=throughput+10*log10 (numel (searchterm)*mytimer)
sk=find (group==class)
[r, c]=size (trainingdata)
Contentavail=sk*c
Responsetime=mytimer
```

Error Value and Accuracy has been calculated using below procedure:

```
Resultterm=sim (net, search_term')
error_value=abs (resultterm-search_value)
Searchaccuracy=100-error_value
```



Figure 5.20: Throughput, Content availability and Response Time by Neural Network

CATEGORY	CLOTHING Y		CLICK HERE TO VIEW ADDED SITES	
		- D ×	ADDED SITES	
SITE NAME ADD CONTENT	Stay in sync with the latest fashion tr in women clothing and look fabulous	rich Accuracy is :98.7722 % Error Percentage is :1.2278	www.amazon.com/clothes www.flipcart	
CLICK HERE	Whether you wish to revamp you wardrobe,	OK		
	add some basic essentials or indulge in an exciting shopping spree, the collection of women's	fashion trends	V	
	19.6429 %Word Processing112/112	VVeight matrix		
	Stay sync with latest	65 61 60 77 72 80 70 88 53 97 ▲ 25 66 104 121 87 61 59 75 79 91 9 52 34 123 33 72 15 91 104 68 108 21 103 82 107 69 77 63	Train Using SVM	
	fashion trends women clothing	47 79 104 66 52 64 103 67 110 103 123 61 56 33 133 49 61 75 117 28 37 37 79 94 54 41 102	Train using LDA	
	REMOVE STOPPING WORDS	ADD THIS SITE		
CLASSIFICATION AND RANKING-				
WHAT ARE YOU LOOKING	FOR	WWW.flipcart	^	
NEURAL	SVM FUZZY LDA	www.amazon.com/clothe	25 🗸	

Figure 5.21: Search Accuracy and Error Percentage score by Neural Network

Figure 5.22: Ranked result on the basis of score of various QoS parameters

As shown in Figure 5.22, after calculations of Quality Factors of different web pages in Figure 5.20 and Figure 5.21, ranking values are to be assigned on the basis of resultant score of quality factors by various classification algorithms.

6.1 RESULTS USING NAÏVE BASED ALGORITHM IN SYMFONY FRAMEWORK

In order to ensure the effectiveness of our search based QoS ranking, we performed experiments on some publicly available web services QoS data which had been collected by Chengying Mao and Jifu Chen [60]. Graphs showing improvement of literature work. In this section we have discussed the proposed model with an example. We assume that user wants to search best jewellery web service from services available in service registry in terms of response time, throughput, recall rate and precision rate.

Services Names	Availability	Response Time
Amazon	0.9	0.76(sec)
Bluestone	0.8	0.91(sec)
Flipkart	0.7	0.41(sec)
Shopclues	0.68	0.98(sec)
Jabong	0.62	0.55(sec)
GehnaJewellers	0.5	0.33(sec)
FabIndia	0.4	0.10(sec)
Voonik	0.3	0.54(sec)
KalyaniJewellers	0.34	0.15(sec)
EverStylish	0.2	0,16(sec)

Table 6.1: QoS Database

The above Table 6.1 shown QoS data base having similar web services with different QoS parameters. Table 6.2 shown the list of web services before applying Ranking algorithm in UDDI registry.

Sr.No	Name of Web Service	Rank		
1	http://www.fabindia.com/women/accessories/jewellery.html	0		
2	http://www.voonik.com/latest/women-jewellery	1		
3	http://shop.mango.com/IN/women/accessories/jewellery			
4	http://www.bluestone.com/	3		
5	http://gehnajewellers.com/	4		
6	http://www.iigj.org/	5		
7	http://everstylish.com/jewelry.html	6		
8	http://www.mjdma.org/	7		
9	http://kalyanjewellers.net/	8		
10	http://www.alibaba.com/premium/jewellery.html?uptime=20141201&ptsid=1	9		
	012000055618814&crea=49000881530&plac=&netw=g&device=c&ptscode			
	=0110202010040001			

Table 6.2: A List of 10 Jewellery Web Services before Rank in UDDI Registry

Below Table 6.3 shown the web services after applying ranking algorithm and provides execution results in terms of QoS factors as well as shown the ranked list of execution results obtained from our model.

Sr.No	Name of Web Service	Rank
1	http://kalyanjewellers.net/	0
2	http://www.alibaba.com/premium/jewellery.html?uptime=20141201&ptsid=1012 000055618814&crea=49000881530&plac=&netw=g&device=c&ptscode=011020	1
	2010040001	
3	http://www.mjdma.org/	2
4	http://www.bluestone.com/	3
5	http://everstylish.com/jewelry.html	4
6	http://www.iigj.org/	5
7	http://shop.mango.com/IN/women/accessories/jewellery	6
8	http://www.fabindia.com/women/accessories/jewellery.html	7
9	http://gehnajewellers.com/	8
10	http://www.voonik.com/latest/women-jewellery	9

Table 6.3: A List of 10 Jewellery Web Services after Rank in UDDI Registry

In Figure 6.1, snapshot of Symfony framework has been shown the various shopping sites to purchase jewellery. Customer can click on individual rank to find out Throughput, Content Availability and Response Time. Overall throughput and response time has been calculated.

Content of HMTL Page Us $ imes$	+					-	٥	×
researchwebtech.com/aman_research/curl/index.php			▼ C ^e Q, Search	☆ 自 ♥	7 +	î	9	-
	Jewenery		Subr	Inrougnput				
		DANK	FEEDBACK					
LINK	febielis	KANK	FEEDDACK					
nttp://www	.tabindia.com/women/accessories/jeweilery.ntml	U	Give Feedback Show Feedback					
http://www	.voonik.com/latest/women-jewellery	1	Give Feedback Show Feedback					
http://shop	.mango.com/IN/women/accessories/jewellery	2	Give Feedback Show Feedback					
http://www	.bluestone.com/	3	Give Feedback Show Feedback					
http://gehr	ajewellers.com/	4	Give Feedback Show Feedback					
http://www	.iigj.org/	5	Give Feedback Show Feedback					
http://ever	stylish.com/jewelry.html	6	Give Feedback Show Feedback					
http://www	.mjdma.org/	7	Give Feedback Show Feedback					
http://kaly	anjewellers.net/	8	Give Feedback Show Feedback					
				Content Based Rank				
								_

Figure 6.1 Comparison of Proposed Model With Existing Model

The web services shown in above Figure 6.1 are not in ranking order imposed by Naïve Baysed Algorithm. Whenever the customer clicks on Content Based Rank button then ranked results will be shown as shown in Figure 6.2.

Get Content of HMTL Page Us × +					-	٥	×
esearchwebtech.com/aman_research/curl/index.php		▼ C Q, Search	☆自て	• •	î	9	≡
							^
jewellery		Subm	Throughput				
							1
LINK	RANK	FEEDBACK					
http://kalyanjewellers.net/	0	Give Feedback Show Feedback					
http://www.mjdma.org/	1	Give Feedback Show Feedback					
http://www.bluestone.com/	2	Give Feedback Show Feedback					1
http://everstylish.com/jewelry.html	3	Give Feedback Show Feedback					
http://www.iigj.org/	4	Give Feedback Show Feedback					1
http://shop.mango.com/IN/women/accessories/jewellery	5	Give Feedback Show Feedback					1
http://gehnajewellers.com/	6	Give Feedback Show Feedback					1
http://www.voonik.com/latest/women-jewellery	7	Give Feedback Show Feedback					
http://www.fabindia.com/women/accessories/jewellery.html	8	Give Feedback Show Feedback					
			Content Based Rank				ļ
					la.	08:08	~
			^ =		⁷⁷ 03	3-10-20	

Figure 6.2 Snapshot of similar web services after ranking in Symfony Framework

🔶 🖲 rese	archwebtech.com/aman_research/cur	l/index.php	▼ Google	Q ☆ 自 🕴	俞
		Analysis Factor	×		
		Over All Response Time: 0.65			
	jewellery	Throughput: 0.5454545454545455		Submit Factor	
	LINK		ANK	FEEDBACK	
	http://kalyanjewellers.net/		0	Give Feedback Show Feedback	
	http://www.alibaba.com/premium/jewe device=c&ptscode=011020201004000	illery.html?uptime=20141201&ptsid=1012000055618814&crea=49000881530&plac=&netw=g D1	& 1	Give Feedback Show Feedback	
	http://www.mjdma.org/		2	Give Feedback Show Feedback	
	http://www.bluestone.com/		3	Give Feedback Show Feedback	
	http://everstylish.com/jewelry.html		4	Give Feedback	

In above Figure 6.2, web services results are to be shown after imposed rank algorithm

Figure 6.3 Snapshot shown Overall Response Time and Throughput of 10 Jewellery Web Services

Above Figure 6.3 shown overall score of quality factors i.e. Response Time and Throughput.

Get Content of HMTL Page Us $ imes$ +				-	- 0	×
(i researchwebtech.com/aman_research/curl/index.php#	c	Q, Search	☆ 自 ♥ ↓	r 9	NEW	≡
Search Search	All <i>ive</i> Convert 👫 Convert Files 🥘 Measurement Converter 🛐 Transla	ıte 🛐 Facebook 🌅 79°F Solan, India			æ	*•
	Factor	×				
	Content Availability: 27					
jeweilery	Response Time:		Submit Factor			
LINK		ANK	FEEDBACK			
http://www.ernestjones.co.uk/webstore/jewe	e Throughput:		Give Feedback			
	0.69		Show Feedback			
http://www.alibaba.com/premium/jewellery.ht	n		Give Feedback			
device=c&ptscode=0110202010040001			Show Feedback			
http://www.mjdma.org/		2	Give Feedback			
			Show Feedback			
🤌 🗿 🚞 🔹 🔌	e	" "	* 2 1 * 1 = 1		01:06 23-06-2	5 016

Figure 6.4 Snapshot shown Individual Quality Factor Scores of 10 Jewellery Web Services

Figure 6.4 shown scores of quality factors of individual web service after clicking on Rank value which is located in correspondence to every web service.

6.1.1 Performance Analysis Using Naïve Baysed Classifier

The proposed approach empirical results are shown through the graphical representations. It shows the proposed system's performance [57] is better than existing approaches.

Parameters Response Time(RT) Throughput(TR) Algorithms d=10 d=30 d=50 d=10 d=30 d=50 Greedy 0.6412 0.7182 0.7877 65.51 77.06 81.85 CloudRank 0.6598 0.7682 0.8336 67.18 80.24 85.67 **PSORank** 0.6728 0.7753 0.8408 80.54 67.38 86.63 **Naïve Baysed** 0.4522 0.6588 0.5856 120.8 99.5 76.69

Table 6.4: Experimental results for algorithms Greedy, CloudRank, PSORank [57] and Naïve Bayes

To facilitate presentation, we denote our proposed ranking algorithm as Naïve Baysed in above Table 6.4. We focus our comparison on the differences between existing algorithms. Comparisons of existing algorithms with proposed algorithm are made with 10 to 70 web services as shown in above Table 6.4 where d shows the number of web services in terms of Response Time(RT) and Throughput(TR).



Figure 6.5 Comparison of Response Time value between Existing and Proposed Method

From the above graph in Figure 6.5, the deviation is shown in proposed method as well as in existing approaches. Research value in proposed approach of Response Time is 0.45 and value in existing approach of Response Time is 0.67. This deviation of response time shows the positive feedback of proposed work in terms of 10 web services.



Figure 6.6 Comparison of Throughput value between existing and proposed method

The graph in above Figure 6.6 shows the throughput value deviations in both approaches. It is clear from the graph that the research value of throughput is greater than the existing value of throughput. Thus it is easier for the user to select the right service using proposed method than the existing method.

Table 6.5	Result	Analysis	of	accuracy	evaluation
-----------	--------	----------	----	----------	------------

70 Services Sample	Existing Values	Proposed Values
RT(Recall)	0.796	0.98
RT(Precision)	0.785	0.96

The above Table 6.5 shows comparison analysis of values derived from proposed methods with the values obtained from existing methods proposed by Michal Athanasopoulos and Kostas Kontogiannis[60].Comparison Analysis shown improvements in terms of accuracy parameters.



Figure 6.7 Comparison of precision value between existing and proposed approaches

The graph in above Figure 6.7 shown the precision deviations in both existing and proposed approaches.Research Value of precision is higher than existing Value of Michal [60].According to Wikipedia, high precision means that an algorithm returned substantially more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results.



Figure 6.8 Comparison of Recall value between existing and proposed approaches

The graph in above Figure 6.7 and Figure 6.8 shown the precision and recall deviations in both existing and proposed approaches. According to Wikipedia, high precision means that an algorithm returned substantially more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results. In both graphs, X axis (horizontal line) shows number of URLs and Y axis shows (vertical line) shows recall and precision range. Both precision and recall are returning positive results as compare to existing approach. As we already discussed in Chapter -4, Naïve Bayes gives poor performance whenever data becomes large.

6.2 Results using other Data Classification Algorithms in MATLAB

To overcome the problems of Naïve Baysed Algorithm, other data classification algorithm has been implemented on E-Commerce web applications, compared results of these algorithms and found the best one.

Various Data Classification Algorithms are under as:

- a. SVM(Support Vector Machine)
- b. Fuzzy Logic
- c. LDA(Linear Discriminant Analysis)
- d. NN(Neural Network)

Analysis has been done on the basis of following parameters:

6.2.1 Results and Analysis using Recall Rate and Precision Rate in MATLAB

6.2.1.1 Recall Rate and Precision Rate using SVM (Support Vector Machine)

6.2.1.2 Recall Rate and Precision Rate using Fuzzy Logic

6.2.1.3 Recall Rate and Precision Rate using LDA (Linear Discriminant Analysis)

6.2.1.4 Recall Rate and Precision Rate using NN (Neural Network)

6.2.1 Results and Analysis using Recall Rate and Precision Rate in MATLAB

Recall Rate and Precision Rate are accuracy parameters. While calculating results, recall rate is higher than precision rate in some situations or recall rate is lower than precision rate in other situations. In this research, we attempted both recall rate and precision rate on higher sides. Here we calculated these accuracy parameters using SVM, Fuzzy, LDA and Neural Network and compared accuracy results of these algorithms and found the best one.

6.2.1.1 Recall Rate and Precision Rate using SVM (Support Vector Machine)

The table 6.6 shows the recall rate and precision rate value based on SVM classifier and the obtained values are Recall rate = 91.271, 92.089, 92.521, 91.898, 91.232, 91.037, 91.2241, 92.0054, 91.561 and 91.890 and Precision rate = 91.003, 91.24, 91.22, 91.34, 91.56, 91.46, 91.25, 91.29, 91.31 and 91.51 respectively.

Sr.No	Recall Rate (%)	Precision Rate (%)
1.	91.271	91.003
2.	92.089	91.24
3.	92.521	91.22
4.	91.898	91.34
5.	91.232	91.56
6.	91.037	91.46
7.	91.2241	91.25
8.	92.0054	91.29
9.	91.561	91.31
10.	91.890	91.51

Table 6.6 Recall and Precision rate values w.r.t. SVM



Figure 6.9 Recall and Precision Rate Graph

The above Figure 6.9 shows the comparable values of the graph using SVM on the basis of Recall rate and Precision rate. It has been seen that Recall rate value is higher than the precision rate value in terms of percentage and different values is around .876 %.

6.2.1.2 Recall Rate and Precision Rate using Fuzzy Logic

Recall rate and precision rate values are shown in the below table 6.7 and the obtained values of Recall rate =94.371, 95.589, 94.421, 94.898, 94.632, 95.937, 94.2241, 94.3054 94.661 and 94.790. And of Precision rate = 94.003, 94.124, 94.122, 94.134, 95.156, 94.146, 94.125, 94.129, 94.131 and 94.151.

Sr.No	Recall rate (%)	Precision Rate (%)
1.	94.371	94.003
2.	95.589	94.124
3.	94.421	94.122
4.	94.898	94.134
5.	94.632	95.156
6.	95.937	94.146
7.	94.2241	94.125
8.	94.3054	94.129
9.	94.661	94.131
10.	94,790	94.151

Table 6.7 Recall and Precision rate values w.r.t. Fuzzy Logic





In the above graph in Figure 6.10, Fuzzy Logic based comparison has been made using two parameters, i.e. Recall rate and Precision rate. It has been noticed that Fuzzy Logic Method has intermediate recall rate and the precision rate in the proposed work w.r.t SVM.

6.2.1.3 Recall Rate and Precision Rate using LDA (Linear Discriminant Analysis)

Recall rate and precision rate values are shown in the below table 6.8 and the obtained values of Recall rate =90.371, 90.589, 90.421, 90.898, 90.632, 90.937, 90.2241, 90.3054, 90.661 and 90.790. And of Precision rate = 90.003, 90.124, 90.122, 90.134, 90.156, 90.146, 90.125, 90.129, 90.131 and 90.151.
Sr.No	Recall rate (%)	Precision Rate (%)
1.	90.371	90.003
2.	90.589	90.124
3.	90.421	90.122
4.	90.898	90.134
5.	90.632	90.156
6.	90.937	90.146
7.	90.2241	90.125
8.	90.3054	90.129
9.	90.661	90.131
10.	90.790	90.151

Table 6.8 Recall and Precision rate values w.r.t. LDA



Figure 6.11. Recall and Precision Rate Graph

In the above graph in Figure 6.11, comparison has been made using two parameters, i.e. Recall rate and Precision rate. It has been noticed that LDA Method has intermediate recall rate and the precision rate in the proposed work w.r.t Fuzzy logic, SVM and neural network classifier.

6.2.1.4 Recall Rate and Precision Rate using NN (Neural Network)

From the result evaluation as shown in table 6.9, it has been seen that the recall rate and precision rate value based on NN classifier came out to be Recall rate = 98.271, 98.089, 99.521, 98.898, 98.232, 98.037, 98.2241, 98.0054, 99.561 and 98.890. Precision rate = 99.003, 98.24, 98.22, 98.34, 98.56, 98.46, 98.25, 98.29, 98.31 and 98.51 respectively.

Sr.No	Recall rate (%)	Precision Rate (%)
1.	98.271	99.003
2.	98.089	98.24
3.	99.521	98.22
4.	98.898	98.34
5.	98.232	98.56
6.	98.037	98.46
7.	98.2241	98.25
8.	98.0054	98.29
9.	99.561	98.31
10.	98.890	98.51

Table 6.9 : Recall and Precision rate values w.r.t. NN



Figure 6.12: Recall and Precision Rate Graph using NN

Above graph in Figure 6.12 describes the comparison values graph using NN on the basis of Recall rate and Precision rate. It has been noticed that Neural Network Method has highest recall rate and the precision rate in the proposed work. As neural network has a good learning rate so it has a better recall rate and precision rate having values 99 and 98 respectively.

Technique	Precision rate	Recall
		rate
Michael et.al [60]	78%	79%
Proposed Work with NN	99%	98%

From the Table 6.10, it has been noted that the Precision rate and Recall rate of the proposed algorithm is good than that of traditional method and we can conclude that the proposed method is able to generate an acceptable result.

6.2.2 Results and Analysis using Throughput and Response Time in MATLAB

Throughput means number of resulting web services with in a period of time and Response Time means how much time a web service requires to open. Here various experiments in MATLAB has been made in terms of throughput and response time using different classification algorithms.

6.2.2.1Throughput and Response Time using SVM (Support Vector Machine)

The table 6.11 represents the comparison between proposed work and existing work with Suport Vector Machine (SVM). Throughput in percentage and response time in seconds and proposed throughput with SVM is better than existing but less as compare to the proposed neural network method.

Throughput (%)		Response Time (sec)	
Existing Work	Proposed Work	Existing Work	Proposed Work
81.83	87.76	0.8417	0.89
85.67	87.21	0.8282	0.91
86.03	85.29	0.8336	0.74
85.36	88.13	0.7877	0.62
86.08	87.39	0.7763	0.88
80.63	84.51	0.7632	0.72
80.19	83.73	0.7753	0.62
80.50	87.31	0.7662	0.83
80.24	86.92	0.7182	0.78
77.06	87.51	0.6748	0.64

Table 6.11: SVM Comparison



Figure 6.13: Throughput Comparison Graph w.r.t. SVM



Figure 6.14: Response Time Comparison Graph w.r.t. SVM

As shown in above graphs in Figure 6.13 and Figure 6.14, our proposed work gives improved results in terms of response time and throughput. But the problem of SVM is that it gives poor performance if number of data items become increase.

6.2.2.2 Throughput and Response Time using Fuzzy Logic

The below table 6.12 represents the comparison between proposed work and existing work with Fuzzy Logic. Throughput in percentage and response time in seconds and proposed throughput with fuzzy logic is better than existing but less as compare to the proposed neural network method and proposed SVM method.

Throughput (%)		Response Time (sec)	
Existing Work	Proposed Work	Existing Work	Proposed Work
81.83	89.17	0.8417	0.62
85.67	89.38	0.8282	0.12
86.03	88.31	0.8336	0.43
85.36	89.52	0.7877	0.71
86.08	80.15	0.7763	0.61
80.63	82.31	0.7632	0.59
80.19	87.14	0.7753	0.54
80.50	88.53	0.7662	0.63
80.24	89.74	0.7182	0.67
77.06	87.93	0.6748	0.61

Table 6.12: Fuzzy Comparison



Figure 6.15: Throughput Comparison Graph w.r.t. Fuzzy



Figure 6.16: Response Time Comparison Graph w.r.t. Fuzzy

As shown in above graphs in Figure 6.15 and Figure 6.16, our proposed work gives improved results in terms of response time and throughput.

6.2.2.3 Throughput and Response Time using LDA (Linear Discriminant Analysis)

The table 6.13 represents the comparison between proposed work and existing work with LDA. Throughput in percentage and response time in seconds and proposed throughput with LDA is better than existing but less as compare to the proposed fuzzy logic method, proposed neural network method and proposed SVM method.

Throughput (%)		Response Time (sec)	
Existing Work	Proposed Work	Existing Work	Proposed Work
81.83	80.03	0.8417	0.519
85.67	89.91	0.8282	0.332
86.03	83.42	0.8336	0.713
85.36	81.43	0.7877	0.639
86.08	88.41	0.7763	0.517
80.63	82.93	0.7632	0.514
80.19	89.39	0.7753	0.321
80.50	88.51	0.7662	0.817
80.24	89.38	0.7182	0.427
77.06	87.94	0.6748	0.315

Table 6.13: LDA Comparison



Figure 6.17 : Throughput Comparison Graph w.r.t. LDA



Figure 6.18: Response Time Comparison Graph w.r.t. LDA

As shown in above graphs in Figure 6.17 and Figure 6.18, our proposed work gives improved results in terms of response time and throughput.

6.2.2.4 Throughput and Response Time using NN (Neural Network)

The below table 6.14 represents the comparison between proposed work and existing work with neural network. Throughput in percentage and response time in seconds and proposed throughput is better than existing work.

Throughput (%)		Response Time (sec)	
Existing Work	Proposed Work	Existing Work	Proposed Work
81.83	93.73	0.8417	0.344
85.67	90.84	0.8282	0.219
86.03	97.69	0.8336	0.413
85.36	94.53	0.7877	0.341
86.08	92.78	0.7763	0.325
80.63	91.72	0.7632	0.419
80.19	96.79	0.7753	0.298
80.50	99.21	0.7662	0.447
80.24	93.42	0.7182	0.349
77.06	95.19	0.6748	0.314

Table 6.14: Neural Network Comparison of Throughput and Response Time



Figure 6.19: Throughput Comparison Graph w.r.t. NN



Figure 6.20: Response Time Comparison Graph w.r.t. NN

As shown in above graphs in Figure 6.19 and Figure 6.20, our proposed work gives improved results in terms of response time and throughput.

6.3 CONCLUSION OF RESULT ANALYSIS

Various classification techniques have been compared on the basis of throughput, response time, recall rate and precision rate for E-Commerce usage. From the result analysis it has been seen that recall rate and precision rate has best value using Neural Network and has worst values using SVM method. Neural Network found to be the best one as compared to another algorithms in terms of Response Time and Throughput values.

6.2 COMPARISION

Issues	Literature Review	Proposed Model
Protocol Comparison	SOAP	REST
Registry Issues	UDDI,WSDL	RDF,JSON
Security Issue	XML	-JSON -Symfony Framework
Model Issue	-Non Adaptable	Adaptable
Algorithm Improvements	Less Efficient	Improved
Optimization Issue	Less Optimized	More optimized

Table 6.15: Comparison of Proposed Model with Existing Model

- Protocol Comparison-SOAP protocol has been used in existing work. Code used in SOAP is very plumbing.
- 2. **Registry Issues** In existing work, UDDI Registry has been used for storage database of similar services.UDDI registry supports only syntactic data, not semantic data. To solve this problem, authors proposed tModel (technical Model) to support semantic data. But the issue in tModel is that it is not secure due to CategoryBag used in tModel.

- 3. **Security Issue**-Anyone by imposing SQLInjection () can break security. But using Symfony Framework, it become more secure.
- 4. **Model Issue-** User can make changes dynamically using Symfony framework. Any new entity can be inserted, removed or modified according to customer's requirements.
- 5. Algorithm Improvement-Naïve Bayes classifier has been implemented using Symfony framework and removed protocol limitations and another classification algorithms like Neural Network, SVM, LDA and Fuzzy Logic has been implemented using MATLAB. Advantages of Neural Network over other networks are that it can work for multiple groups in single iteration. Artificial Intelligence features can be introduced with the help of Neural Classifier.
- 6. **Optimization Issue** Every time at the time of fetching data, data is scrapped and compared dynamically using ranking algorithm.

7.1 SUMMARY

Web services guarantee Web Service providers and clients that are concerned with the QoS. The new discovery approach is suggested to be the solution for the current web service discovery problems which are discussed in Chapter-2.Many E-Commerce sites has been taken to measure the quality factor of things available in different shopping sites and various tools has been applied which are mentioned in Chapter-3. In Chapter-4, more semantic interpretability of data has been provided with the help of RDF. To solve the problems in communication protocols i.e. UDDI and WSDL which represent only syntactic information, tModel was introduced for semantic handling which remained unsecure. REST protocol removes the problem of SOAP protocol in terms of security and storage issue. In this research, RDF and JSON formats overcomes the problem of tModel. RDF and JSON formats are to be used under Symfony Framework which provides more adaptability, reusability and security. Advertised information by service provider is expressed in and is stored and updated using RDF formats in UDDI registry. As mentioned in Chapter-4 and Chapter-5, data values from different sites has been fetched, compared, ranked and calculated quality score as well as assigned to different sites returning similar results using Symfony and MATLAB tools. Finally, buyer can select a shopping page by seeing highest quality score as well as according to the highest rank. Highest QoS (Quality Score) means Highest Rank Composition of REST and RDF .In Chapter -6, various classification techniques have been compared on the basis of throughput, response time, recall rate and precision rate for E-Commerce usage. From the result analysis it has been seen that recall rate and precision rate has best value using Neural Network and has worst values using SVM method.

7.2 OUTLOOK

New Suggested approach with MATLAB gives a complete dialog of the all the stages in Web services and related works in this field. Web services puts a push to focus important data from the optional information acquired from the correspondences of the clients with the Web. Beyond E-marketing, proposed model would be able to work on other services like Blog fetch, Newsfeed fetch, Stock exchange Market as well as travelling agency for airline fares. It will become revolutionary if other search engine provides results after sensing input using neural approach. Further, hybridization of Neuro and Fuzzy can be implemented as per the

increasing requirement of adaptive intelligent system to explain the problems of real world. In future, MATLAB Server can be advantageous with REST protocols benefits.

REFERENCES

- [1] 3rd International Conference on Web Services Orlando, Florida (USA), July 11.
- [2] J. Clerk Maxwell, "A Treatise on Electricity and Magnetism", 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68– 73.
- [3] M. N. Huhns," Agents as Web Services", Internet Computing, IEEE,6(4):93-95,2002.
- [4] Liliana Cabral1, John Domingue1, Enrico Motta1, Terry Payne2 and Farshad Hakimpour1,"Approaches to Semantic Web Services: Overview and Comparisons" Published in The Open University's repository of research publications and other research outputs, UK.
- [5] R. Jennings, "On Agent Based Software Engineering, "Artificial Intelligence,(117):277-296,2000
- [6] Daniel et al., "Improving Text Search Process using Text Document Clustering Approach", International Journal of Science and Research Cohen, W.J., Singer, Y.: Context-sensitive learning methods for text categorization. In: SIGIR 1996: Proc. 19th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 307–315.
- [7] Valentine Janev, Sanja Vranes, "Semantic Web Technologies: Ready for Adoption," IASTED Journal of Robotics and Automation, 12(4), pp.135-145.
- [8] Han, R., et al., "Web Splitter: a unified XML framework for multi-device collaborative Web browsing", In Proceedings of the 2000 ACM conference on Computer supported cooperative work, pp. 221-230.
- [9] J. Cardoso, "Quality of Service and Semantic Composition of Work Flows" PhD Thesis, University of Georgia, 2002.
- [10] E.M Maximilien, M.P Singh, Conceptual Model of Web Service Reputation .ACMSIGMOD Record,2002,31(4),pp.36-41.
- [11] P. Farkas and H. Charaf, "Web Services Planning Concepts", 1st International Workshop on C# and .NET Technologies on Algorithms, Computer Graphics, Visualization, Distributed and WEB Computing, Feb. 2003.
- [12] A.ShaikAli, O.F.Rana, R.Al-Ali and D.W.Walker, "UDDIe: An extended registry for web services". In Proc. Of the Symposium on Applications and the Internet workshops, IEEE CS, pp 85-89, 2003.
- [13] Octavian and S. Jeffrey, "B kd-Tree: A Dynamic Scalable kd-Tree", Department of Computer Science, Duke University Durham, NC 27708, USA.
- [14] S. Kalepu, S. Krishnaswamy and S. W. Loke, "Verity: a QoS metric for selecting Web services and providers," Web Information Systems Engineering Workshops, 2003. Proceedings. Fourth International Conference on, pp. 131-139.
- [15] Hongan Chen, Tao Yu, Kwei-Jay Lin, "QCWS: an implementation of QoS-capable multimedia web services", IEEE Fifth International Symposium on Multimedia Software Engineering, IEEE 2003.
- [16] E.M Maximilien and M.P. Singh,"A Framework and Ontologyfor Dynamic Web Services Selection", IEEE Internet Computing ,vol.8(5),pp:84-93,2004
- [17] S.Ran, A Model for Web Services Discovery with QoS. SIGEcom Exchanges, Vol. 4(1), pp.1–10, 2004
- [18] S. Kalepu, S. Krishnaswamy, S. W. Loke, "Reputation = f(User Ranking, Compliance, Verity)," icws, p. 2004, IEEE International Conference on Web Services (ICWS'04), 2004
- [19] L. Zeng, B. Benatallah, A. H. H. Ngu, M. Dumas, J. Kalagnanam and H. Chang, "QoS-Aware Middleware for Web Services Composition," IEEE Transactions on Software Engineering, vol. 30, no. 5, pp. 311-327, May, 2004.
- [20] Liu, Y., Ngu, A., and Zeng, "QoS Computation and Policing in Dynamic Web Service Selection", Proceedings of the 13th International World Wide Web Conference.
- [21] Changing Sun, et al., "Comparison of UDDI Registry Replication Strategies", Proceedings of the IEEE International Conference on Web Services (ICWS'04) 0-7695-2167-3/04 \$ 20.00 IEEE.
- [22] Werner, et al., "Compressing SOAP Messages by using Differential Encoding", Proceedings of IEEE International Conference on Web Services (ICWS'04), July 6-9, San Diego, CA, USA, pp. 540-547.
- [23] Jia Zhang, et al., "A Service-Oriented Multimedia Componentization Model", International Journal of Web Services Research, 2 (1), pp.54-76.
- [24] M.A. Serhani ,R.Dssouli,A.Hafid and H.Sahraoui ,"A QoS broker based architecture for efficient web services selection ". In *Proc. of the IEEE Int'I Conf. on Web Services ,IEEE CS,2005.*
- [25] L-. H. Vu, M. Hauswirth, and K. Aberer: QoS-based service selection and ranking with trust and reputation
- management, Proceedings of OTM'05, R. Meersman and Z. Tari (Eds.), LNCS 3760, p.p. 466-483, 2005.
- [26] Rina Panigrahy, "Nearest Neighbour Search using Kd-trees" citeseerx.ist.psu.edu.2006.
- [27] Chia-Hui, K. Mohammed, R. Moheb and S. Khaled, "A Survey of WebInformationExtractionSystems", IEEE Transactions on Knowledge and DataEngineering, Tkde-0475-1104.R3.Extraction
- [28] Le Hung et al., "Towards P2P-Based Semantic Web Service Discovery with QoS Support", Business Process Management Workshops, Volume 3812 of the series Lecture Notes in Computer Science pp 18-31.
- [29] Ziqiang Xu, Patrick Martin, Wendy Powley and Farhana Zulkernine, (2007). "Reputation-Enhanced QoS-based Web Services Discovery". IEEE International Conference on Web Services (ICWS 2007).
- [30] Jing Ni, et al., A Semantic Web Service-Oriented Model for E-Commerce. 1-4244-0885-7/07/\$20.00 © IEEE.
- [31] M.Jose, L.Javier, "A Tool for Web Usage Mining", 8th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'07), 16-19 December.

- [32] Liu H., Keselj V., "Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users 'future requests", Data & Knowledge Engineering, 61 (2), pp. 304–330.
- [33] Yang S.Y., Chen, P.Z. and Sun, C.H., "Using Petri Nets to Enhance Web Usage Mining, Acta Polytechnic a Hung Arica", 4 (3), pp. 113-125.
- [34] Yu Liu, "Networked Monitoring System Based on Embedded Dynamic Web", Fuzzy Information and Engineering, Volume 54 of the series Advances in Soft Computing pp 575-584.
- [35] H.Alani et al., "Building a Pragmatic Semantic Web", IEEE Intelligent Systems, pp.61-68.
- [36] Mohamed Adel Serhani1, Elarbi Badidi1, Abdelghani Benharref, Mohamed Salem, "A cooperative approach for QoS-aware Web services' Selection", Proceedings of the International Conference on Computer and Communication Engineering 2008 May 13-15, Kuala Lumpur, Malaysia.
- [37] N. Olfa, S. Maha, S. Esin, B. Antonio and G. Richard, "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites, IEEE Transactions On Knowledge And Data Engineering, 20 (2).
- [38] T. Rajendran and Dr. P. Balasubramanie, "Analysis on the Study of QoS-Aware Web Services Discovery", Journal of Computing <u>http://sites.google.com/site/journalofcomputing/</u>.
- [39] Dou, Z., et al., "Evaluating the Effectiveness of Personalized Web Search". IEEE Transactions on Knowledge and Data Engineering, 21(8), pp: 1178-1190.
- [40] Pak, Alexander, and Patrick Paroubek,"Twitter as a Corpus for Sentiment Analysis and Opinion mining of big data." LREC, 10.
- [41] Sujatha V., Punithavalli, "An Approach To User Navigation Pattern Based On Ant Based Clustering And Classification Using Decision Tress", International Journal Of Advanced Engineering Sciences And Technologies, 1(2), pp. 112 – 117.
- [42] Agarwal, Apoorv, et al."Sentiment analysis of twitter data." Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics,
- [43] Netra Patil and Dr. Arpita Gopal, "Comparative Study of Mechanisms for Web Service Discovery based on Centralized Approach Focusing on UDDI. International Journal of Computer Applications 14(1), pp. 28–31.
- [44] L.Vijayashri and J. Madhuri, "Data Preprocessing in Web Usage Mining", International Conference on Artificial Intelligence and Embedded Systems (ICAIES'2012) July 15-16.
- [45] B.Geeta, T.Shashikumar, R.Prasad, "Literature Survey on Web Mining", IOSRJCE, ISSN: 2278-0661, ISBN: 2278-8727, 5(4), pp. 31-36.
- [46] P.Nithya and P.Sumathi, "A Survey on Web Usage Mining: Theory and Applications", Int.J.Computer Technology & Applications, ISSN: 2229-6093, Vol 3 (4), pp. 1625-1629.
- [47] Jebaseeli, A. Nisha, and E. Kirubakaran, "A Survey on Sentiment Analysis of (Product) Reviews." International Journal of Computer Applications 47.11.
- [48] Shariq Hussain, et al., "Survey on Services Composition Synthesis Model", International Journal of Computer Science Issues, 10 (1).
- [49] Ch Ram Mohan Reddy, and RV Raghavendra Rao, "QOS of Web Service", Survey on Performance and Scalability. pp. 65–73.
- [50] Anuradha et al., "Neural Network Approach for Text Classification using Relevance Factor as Term Weighing Method", IJCA Journal, 68, pp-37-41.
- [51] Mohammad, "Approximate K-Nearest Neighbour Based Spatial Clustering Using K-D Tree", International Journal of Database Management Systems (IJDMS), 5(1).
- [52] Balahur, Alexandra, et al., "Sentiment analysis in the news." arXiv preprint.
- [53] Gupta, Aditi, et al., "Sentiment analysis for social media." International Journal of Advanced Research in Computer Science and Software Engineering 3(7), pp. 216-221.
- [54] Yu, Dongjin, et al., "Personalized QoS prediction for Web services using latent factor models". Services Computing (SCC), 2014 IEEE International Conference on. IEEE
- [55] V.Deepika, K. Namita and M. Neha, "Comparison of Brute-Force and K-D Tree Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, 3(1)
- [56] Ramakrishnan M. et al., "Modified K-Means Algorithm for Effective Clustering of Categorical Data Sets", International Journal of Computer Applications, 89.
- [57] Mao, Chengying, et al., "Search-based QoS ranking prediction for web services in cloud environments". Future Generation Computer Systems, 50, pp. 111-126.
- [58] Muhammad Waqas Khan and Eram Abbasi, "Differentiating Parameters for Selecting Simple Object Access Protocol (SOAP) vs. Representational State Transfer (REST) Based Architecture", Journal of Advances in Computer Networks, 3 (1): pp. 63-66.
- [59]Zainab Aljazzaf, "Bootstrapping quality of Web Services", Journal of King Saud University Computer and Information Sciences, 27(3), pp. 323–333.
- [60] Michael Athanasopoulos and Kostas Kontogiannis, "Extracting REST resource models from procedure-oriented service interfaces", 100, pp.149–166.
- [61] Horakova, Marketa, "Sentiment Analysis Tool using Machine Learning." Global Journal on Technology.
- [62] Omar Adil M.Ali," Comparison between the effects of Different Types of Membership Functions on Fuzzy Logic Controller Performance" in International journal of Emerging Research and Technology Vol3,Issue 3,March 2015.

[63] Huan Xu," Robustness and Regularization of Support Vector Machine" in Journal of Machine Learning Research ,July 2009.