BIG DATA ANALYTICS APPROACH TO MEASURE ACTUAL CUSTOMER BEHAVIOUR

[1]Revati Raman Dewangn, [2]Deepali Thombre, [3]Vivek Parganiha
[1]*Asst. Prof. CSE, CCET, Bhilai,* [2]*Asst. Prof. CSE, SSEC, Bhilai* , [3]*Assoc. Prof, CSE, BIT, Durg*
[1]revati2004@gmail.com, [2]deepthombre@gmail.com , [3]vivekparganiha@gmail.com

***Abstract*— A conventional marketing approach for predicting consumer behaviour is to draw behavioural intention instead of actual behaviour. Frederick F. Reich held, developer of the very popular Net Promoter Score (NPS), based his loyalty system on a customer survey that asks how likely they are to recommend a company's product or service to their friends. Customer satisfaction measurement allows an organisation to understand the issues, or key drivers, that cause satisfaction or dissatisfaction with a service experience. When an organisation is able to understand how satisfied its customers are, and why, it can focus its time and resources more effectively. Measuring customer satisfaction is just one stage in a continuous programme of service transformation.**

For organisations new to this process, the first stages require a review of what the service provides, where it sits in context with other related services in customers' minds, who its customers are and what information about the customer experience is already available. After this, qualitative research should be conducted with customers and staff to highlight key issues that the survey will need to capture. At this point decisions will need to be made about which customers should be interviewed and what methods should be used. Customer satisfaction and measurement issues have very important roles for businesses in providing and maintaining a reasonable advantage. It is recognized that the businesses forming components of marketing mix by acknowledging the customers' prospect, receive customer loyalty and profit in return. Via measuring customer satisfaction, organizations do not only have customer facts also have competitors' knowledge in the market. Big data help to keep track of customer behaviour from huge amount of on line data.

Key Word: Big data, Marketing, Customer behaviour, NPS, Data Base.

I. INTRODUCTION

Customer related measurements such as Net Promoter Score (NPS), Customer Effort Score (CES) and Customer Satisfaction Index (CSI) have all made their way into organisations. The intent is to understand the customer issues and concerns better and "fix" the problem. The painful reality is that while these scores help improve the customer transactions; they do not reveal deep underlying issues that affect customers [1,15].

A sore example is where an organisation uses NPS to measure customer satisfaction in the unlike touch points. The customer gives a positive score after an online sales transaction, but calls the call centre because of a problem with the delivery of the service. The call centre gets a negative score on the sales transaction that actually failed in another channel. The issue here is one of not following customers' behaviours – which are different to customer opinion.

Using this business process, you can determine, for example, whether a customer is likely to terminate their contract[2,3,14]. You can use this prediction as the basis to decide, for example, whether and how customers are to be divided into target groups or whether and how they are to be addressed in a marketing campaign.

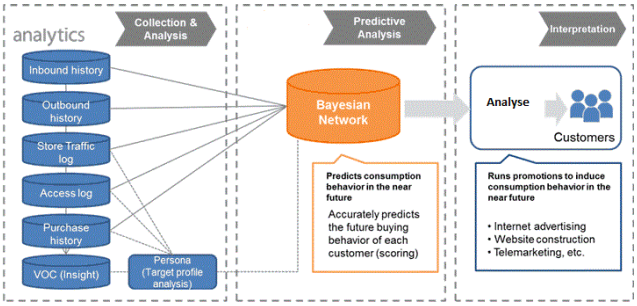

Figure 1 Predict Customer Behaviour

1. System provides customer master data and customer profiles
2. System provides the transaction data: This data such as customers' payment behavior is collected in the operational processes.
3. System provides the transaction data: This data such as activities and sales orders is collected in the operational processes.
4. System updates data in customer knowledge base: By regularly extracting data from SAP CRM and SAP R/3 into SAP BW, the system updates the data in the customer knowledge base in SAP BW. This way, the data can be applied in different analyses and for data mining methods.

5. Create and train model for scoring or decision trees: You create a model for the data mining method Scoring or Decision Tree and train the model using historic data. Both of these data mining methods are used to classify your customers according to known or unknown criteria [17,18].

6. System determines propensities and scores When the model is trained, the system looks for definite patterns in the data. In the case of scoring, for example, you assign to each customer a score that is calculated from specific attributes or behavior attributes and the respective weighting for those attributes. In the case of decision trees, the system determines which customer attributes can be associated with a particular type of customer behavior or propensity.

7. Apply model to new data to predict customer behavior: To predict customer behavior, you apply the model on other data such as data relating to customers or prospects for whom you only have incomplete data.

8. Analyze decision trees or scoring results: In the Analysis Process Designer (APD), you analyze the prediction results.

9. Update customer knowledge base.You update the prediction results in the customer knowledge base[21,22].

10. Transfer customer propensities and scores.

11. System updates the customer behavior information

II. BIG DATA TO PREDICT CUSTOMERS' BEHAVIOURS

Sensible (and amusing) as it sounds, his dictum no longer rings true. The Age of Big Data has arrived and, with it, the ability to predict the future is increasingly a part of a new business reality. Whatever your discipline, doing business today means immersing yourself, and your organization, in a wealth of messy, unstructured, real time data from customers, competitors, and markets and finding ways to use such data visibility to see what's coming[4,26].

Advantage lies in a capacity to predict the future before your rivals can — whether they're companies or criminals. Consider how the New York Police Department is using Big Data to fight crime in Manhattan. According to a series on Big Data in *The New York Times*, the NYPD and other big city police departments are using data-crunching technology to geo-locate and analyze "historical arrest patterns," while cross-tabbing them with sporting events, paydays, rainfall, traffic flows, and Federal holidays to identify what NYPD calls likely crime "hot spots."

As immortalized in a"Smarter Planet" commercial from IBM, such insight can help deploy officers to locations where crimes are likely to occur before they are actually committed[5,6.28].

The beauty of such Big Data applications is that they can process Web-based text, digital images, and online video. They can also glean intelligence from the exploding social media sphere, whether it consists of blogs, chat forums, Twitter trends, or Facebook commentary. Traditional market research generally involves unnatural acts, such as surveys, mall-intercept interviews, and focus groups. Big Data examines what people say about what they have done or will do. That's in addition to tracking what people are *actually* doing about everything from crime to weather to shopping to brands. It is only Big Data's capacity for dealing with vast quantities of real-time unstructured data that makes this possible[7,8,9].

For example, retailers like Wal-Mart and Kohl's are making use of sales, pricing, and economic data, combined with demographic and weather data, to fine-tune merchandising store by store and anticipate appropriate timing of store sales. Similarly, online data services like eHarmony and Match.com are constantly observing activity on their sites to optimize their matching algorithms to predict who will hit it off with whom. The same logic is being applied to economic forecasting.

For example, the number of Google queries about housing and real estate from one quarter to the next turns out to predict more accurately what's going to happen in the housing market than any team of expert real estate forecasters. Similarly, Google search queries on flu symptoms and treatments reveal weeks in advance what flu-related volumes hospital emergency departments can expect[23,24].

Much of the data organizations are crunching is human-generated. But machine sensors — what GE

,

people like CMO Beth Comstock called "machine whispering" when I talked with her this past summer — are creating a second tsunami of data.

Digital sensors on industrial hardware like aircraft engines, electric turbines, automobiles, consumer packaged goods, and shipping crates can communicate "location, movement, vibration, temperature, humidity, and even chemical changes in the air." As the volume of both human and machine data grows exponentially, so too will organizations' ability to see the future.

The net of all this is hardly a cold quantitative world. Rather, as marketers and machine systems learn more about our attitudes and behaviors, they're likely to achieve greater intimacy with consumers and customers than ever before. Yes, there is the risk of an Orwellian nightmare, if the inferences from Big Data become too intimate and too intrusive — and end up in the wrong hands. But there is also the opportunity to deliver services and marketing with unprecedented precision and accuracy, meeting and exceeding customer expectations in preternatural ways at every turn. Knowing the right time to deliver the right message (or action) in the right place *before* the time has come will bestow extraordinary power to those who wield such intelligence with intelligence. Use prediction wisely, and Big Data has the potential to make the world small again[10]. That is every marketer's dream: getting closer to customers.

Understanding which clients are your most precious shoppers is vital because it enables you in addition cognizance your efforts. Historically, groups frequently define their most treasured people (MVPs) because the consumers who spend the maximum cash. But, you can locate these clients are the most high-priced to preserve and are the least loyal over the long time.

Large information comes into play here through calculating several factors that help you get greater records about your customers. Now that you have your personas, let's find out greater about them by using calculating the metrics underneath.

Average buy length: How a good deal do your clients spend on a normal purchase? Examine this no longer simply in mixture, however by using every personality. Additionally think about the fact that humans purchase based on value, now not totally on price. Can you promote greater to any of your personas the usage of promotions to develop attention and hobby in different products?

Lifetime value: How a good deal money does the consumer character spend with you over their lifetime? Is it loads? Or is it a touch? This metric is indicative of the relationship you've got together with your customers.

Acquisition prices: How a great deal have you spent on marketing and sales to get this sort of consumer? In case you spend plenty, permit's wish that your clients do not fee much to keep and they make large purchases from you. If that isn't always the case, you may need to re-examine your acquisition methods.

Retention prices: What do your shoppers want from you so one can live? Do they need a variety of aid, education, or verbal exchange? Normally it costs extra to acquire a customer than keep them. Ensure that you are doing all of your great to construct relationships with your clients and make them sense valued.

Purchaser happiness: Are your clients happy along with your products or services? Are there companies of happy and unhappy clients, and what's the distinction among the 2? Investigating this will display flaws, highlights important enhancements and may even activate you to regulate purchaser expectations.

Fee alignment: Are your supposed customers truely shopping for from you? If the meant middle clients aren't buying from you, then who is? This may assist you to refine your patron personas, especially if it looks like you're out of alignment.

That is where large data analytics comes in to play. Try to tease out demographic and behavior traits that correlate together with your great customers (consumers whose lifetime cost is more than the combination of acquisition and retention fees) and fit them on your personas. Also preserve an eye fixed out for customers who're moderately treasured and people who don't appear to healthy the mould.

The favored stop result is several agencies of customers described through behaviour, demographics and merit. Those businesses must all be prioritised by the value they offer to your corporation.

,

A exceptional result of teasing out behaviour developments is identifying purchasing drivers after which tailoring advertising and marketing contact-factors. Say you've got a fee-conscious patron who abandons the shopping cart. Sending that purchaser a 20% off cut price together with his/her cart gadgets simply might do the trick. For emotionally driven or socially aware clients, a product with proceeds reaping rewards a selected purpose may additionally promote greater than a charge promoting.

III. EXPERIMENT SETUP

For our research we are going to be used test sample data as NYC_Social_Media_Usage.csv by NYC social media web site. It is freely available for test and research. For writing java program we are using notepad++ v6.9., Java development kit version is JDK 1.7 for java environment and hadoop 2.3 for windows, and Windows 8.1 operating system. Here in this web site "http://www.codeproject.com/Articles/757934/Apache-Hadoop-for-Windows-Platform" the installation procedure is given, follow those steps for setting up hadoop environment[11].

Open cmd prompt in admin mode and start hadoop demon using F:\hadoop\Hadoop-2.3-master\Hadoop-2.3-master\sbin\start-yarn command snap shot is follows.

After this start dfs by using F:\hadoop\Hadoop-2.3-master\Hadoop-2.3-master\sbin\start-dfs command the snapshot in figure 4

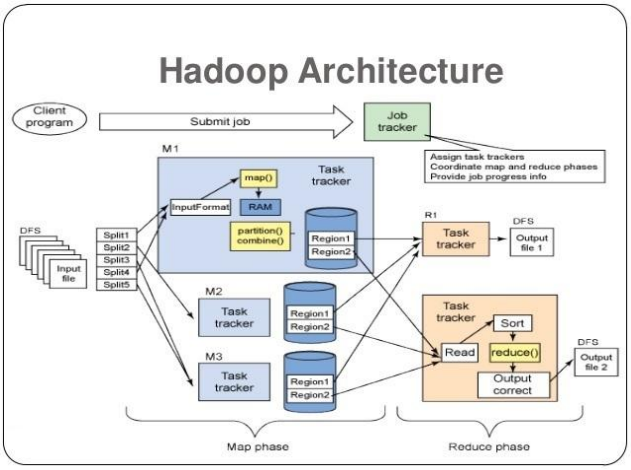

Figure 2 Hadoop Architecture

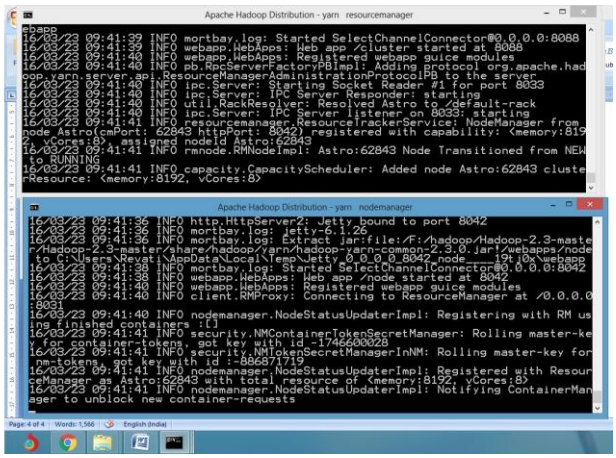

Figure 3 Yarn Hadoop

The skeleton of our research program is given by follow which consist mapper reducer code for our research aim.

```
import java.io.DataInput;
import java.io.DataOutput;
import java.io.File;
import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import
org.apache.hadoop.io.WritableComparable;
```

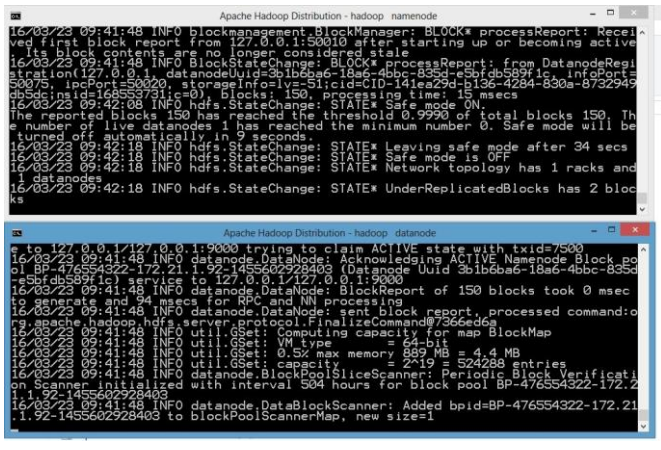

Figure 4 DFS Of Hadoop

```
import org.apache.hadoop.io.WritableUtils;
import org.apache.hadoop.mapreduce.Job;
```

```
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import
org.apache.hadoop.mapreduce.lib.input.FileInputFor
mat;
import
org.apache.hadoop.mapreduce.lib.output.FileOutput
Format;
public class CompositeKeyMR {
public static class CompositeKeyMapper extends
Mapper<LongWritable, Text,
CompositeGroupKey, IntWritable> {
CompositeGroupKey cntry = new
CompositeGroupKey();
Text cntText = new Text();
Text stateText = new Text();
IntWritable populat = new IntWritable();
public void map(LongWritable key, Text value,
Context context)
throws IOException, InterruptedException {
String line = value.toString();
String[] keyvalue = line.split(",");
populat.set(Integer.parseInt(keyvalue[5]));
CompositeGroupKey cntry = new
CompositeGroupKey("",keyvalue[1]);
context.write(cntry, populat);}}
public static class CompositeKeyReducer
extends
Reducer<CompositeGroupKey, IntWritable,
CompositeGroupKey, IntWritable> {
private IntWritable result = new IntWritable();
public void reduce(CompositeGroupKey key,
Iterable<IntWritable> values,
Context context) throws IOException,
InterruptedException {int sum = 0;
for (IntWritable val : values) { sum +=
val.get();
result.set(sum); } context.write(key,
result);}}
private static class CompositeGroupKey
implements
WritableComparable<CompositeGroupKey> {
public CompositeGroupKey() { }
public void write(DataOutput out) throws
IOException {}
public void readFields(DataInput in) throws
IOException
public int compareTo(CompositeGroupKey pop)
{ }
@Override public String toString() { }}
public static void main(String[] args) throws
IOException,
ClassNotFoundException, InterruptedException
{
Configuration conf = new Configuration();
Job job = Job.getInstance(conf,
"CompositeKey");
job.setJarByClass(CompositeKeyMR.class);
job.setMapperClass(CompositeKeyMapper.class)
;
job.setReducerClass(CompositeKeyReducer.clas
s);
job.setOutputKeyClass(CompositeGroupKey.cla
ss);
FileInputFormat.addInputPath(job, new
Path(args[0]));
FileOutputFormat.setOutputPath(job, new
Path(args[1]));
System.exit(job.waitForCompletion(true) ? 0 : 1);
}}
```

IV. EXPERIMENTAL OUTPUT

After running the above code we find the following result given below. According to like was the popularity of web site in table 1.1 and for supporting the graph in figure 5

Table 1.1

Application	Likes
Android	55657
Broadcastr	41306
Facebook	51197
Flickr	52162
Foursquare	63863
Google+	51823
Instagram	22463
Linked-In	50297
Newsletter	51747
Pinterest	16613
SMS	40926
Tumblr	33614
Twitter	57804

,

Vimeo	3471
WordPress	45732
YouTube	64075
iPhone	40450
iPhone App	76736

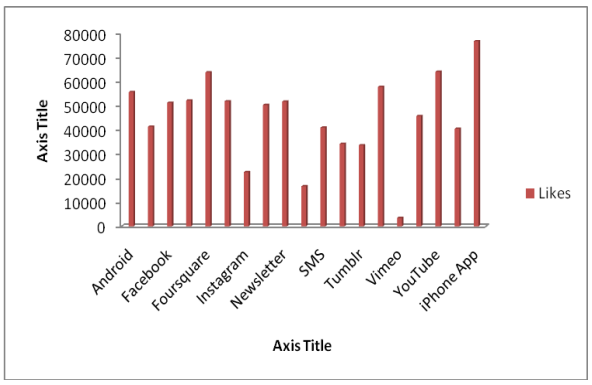

Figure 5 According to likes wise the popularity of web site uses

According to like was the popularity of web site in table 1.2 and for supporting the graph in figure 6

Table 1.2

Applicatio	followers
Android	304476
Broadcastr	190484
Facebook	283377
Flickr	251707
Foursquare	212215
Google+	228409
Instagram	126380
Linked-In	270154
Newsletter	287058
Pinterest	107256
SMS	214266
TOTAL	141174
Tumblr	279238
Twitter	209023
Vimeo	18043
WordPress	264297
YouTube	191429
Youtube	149339
iPhone	204422
iPhone App	459255
nyc.gov	153771

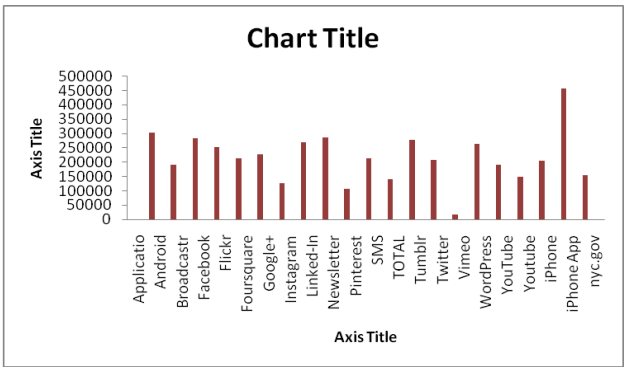

Figure 6 According to followers wise the popularity of web site uses

According to visit wise the popularity of web site in table 1.3 and for supporting the graph in figure 7

Table 1.3

Application	visits
Android	32677
Broadcastr	29699
Facebook	55944
Flickr	46376
Foursquare	48221
Google+	53116
Instagram	22694
Linked-In	61283
Newsletter	52685
Pinterest	22253
SMS	27484
TOTAL	31396
Tumblr	47118
Twitter	74195
Vimeo	9877
WordPress	47103
YouTube	61270
iPhone	50150
iPhone App	56190
nyc.gov	20211

,

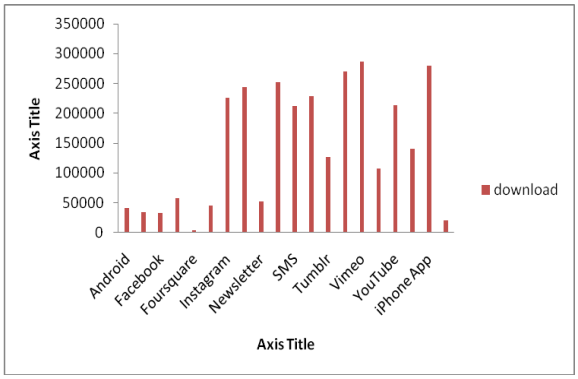

Figure 7 According to visits of customer the popularity of web site uses

According to download wise the popularity of web site in table 1.4 and for supporting the graph in figure 8

Table 1.4

Application	download
Android	40926
Broadcastr	34178
Facebook	33614
Flickr	57804
Foursquare	3471
Google+	45732
Instagram	226223
Linked-In	243213
Newsletter	52685
Pinterest	251707
SMS	212215
TOTAL	228409
Tumblr	126380
Twitter	270154
Vimeo	287058
WordPress	107256
YouTube	214266
iPhone	141174
iPhone App	279238
nyc.gov	20211

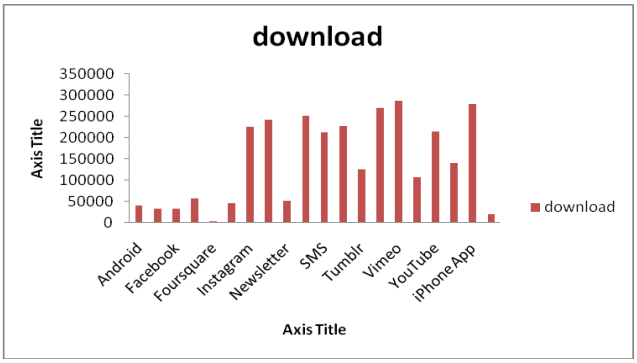

Figure 8 According to downloads of customer the popularity of web site uses

V. CONCLUSION

In this paper, we apply Hadoop for large-scale log analysis and we use NYC_Social_Media_Usage data set as our test data. Our main objective is to efficiently cluster these data by proposing the appropriate number of clusters for particular websites and proper amount of entries using Hadoop framework. With Hadoop and our preprocessing, our system can support large log files. Through our experiments we can deduce the popularity of a web site which have large numbers of likes/visits/followers/download. The experiment will provide sufficient result and give advance result to secure the popularity of different web site. It will be helpful for marketing researchers and advertisement agency for magnify their promotional areas and uplift the marketing strategy.

REFERENCES

[1] apache mahout [Online]. http://en.wikipedia.org/wiki/Apache Mahout.
[2] confusion matrix [Online]. http://en.wikipedia.org/wiki/Confusion matrix.
[3] DaviesBouldin index [Online]. http://en.wikipedia.org/wiki/DaviesBouldin index,.
[4] MapReduce [Online]. http://en.wikipedia.org/wiki/Mapreduce.
[5] overview - apache mahout - apache software foundation [Online]. https://cwiki.apache.org/confluence/display/MAHOUT/Overview.
[6] receiver operating characteristic [Online]. http://en.wikipedia.org/wiki/Receiver operating characteristic.
[7] sensitivity and specificity [Online]. http://en.wikipedia.org/wiki/Sensitivity and specificity.
[8] silhouette (clustering) - wikipedia, the free encyclopedia [Online]. http://en.wikipedia.org/wiki/Silhouette (clustering).
[9] welcome to apache hadoop! [Online]. http://hadoop.apache.org/.
[10] A. Chandrasekhar and K. Raghuveer. Intrusion detection technique by using k-means, fuzzy neural network and svm classifiers. In *Computer Communication and Informatics (ICCCI), 2013 International Conference on*, pages 1–7, 2013.

,

[11] D. Denatious and A. John. Survey on data mining techniques to enhanceintrusion detection. In *Computer Communication and Informatics (ICCCI), 2012 International Conference on*, pages 1 –5, jan. 2012.
[12] T. Fawcett. An introduction to roc analysis. *Pattern Recogn. Lett.*, 27(8):861–874, June 2006.
[13] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. Clustering algorithms and validity measures. In *Scientific and Statistical Database Management, 2001. SSDBM 2001. Proceedings. Thirteenth InternationalConference on*, pages 3 –22, 2001.
[14] R. Naidu and P. Avadhani. A comparison of data mining techniques for intrusion detection. In *Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on*, pages 41–44, 2012.
[15] K. M. Passino, Fuzzy Control, Addison-Wesley, 1998.
[16] E. H. Mamdani, S. Assilian, *An experiment in linguistic synthesis with a fuzzy logic controller*, International journal of man-machine studies 7 (1) (1975) 1–13.
[17] T. Takagi, M. Sugeno, *Fuzzy identification of systems and its applications to modeling and control*,*Systems, Man and Cybernetics*, IEEE Transactions on SMC-15 (1) (1985) 116–132.
[18] Michaelis, C.D., Ames, D.P., Apr. 2009. *Evaluation and implementation of the OGC web processing service* for use in client-side GIS. GeoInformatica 13 (1), 109e120.
[19] Michener, W., Vieglais, D., Vision, T., Kunze, J., Cruse, P., Jan_ee, G., Jan. 2011. *DataONE: data observation network for earth e preserving data and enabling innovation in the biological and environmental sciences*. D-Lib Mag. 17 (1/2).
[20] Nativi, S., Caron, J., Davis, E., Domenico, B., Nov. 2005. Design and implementation of netCDF markup language (NcML) and its GML-based extension (NcML-GML). Comput. Geosci. 31 (9), 1104e1118.
[21] Nielsen, M., 2011. Reinventing Discovery: the New Era of Networked Science. Princeton University Press.
[22] Niu, X., Lehnert, K.A., Williams, J., Brantley, S.L., Jun. 2011. CZChemDB and Earth-Chem: advancing management and access of critical zone geochemical data. Appl. Geochem. 26, S108eS111.
[23] Overpeck, J.T., Meehl, G.A., Bony, S., Easterling, D.R., Feb. 2011. Climate data challenges in the 21st century. Science (New York, N.Y.) 331 (6018), 700e702.
[24] O'Sullivan, B., 2009. Mercurial: the Definitive Guide. In: Definitive Guide Series, vol. 7. O'Reilly Media, Inc.
[25] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, others, DBpedia–A large-scale, multilingual knowledge base extracted from Wikipedia, Semantic Web.
[26] D. Dubois, H. Prade, *The three semantics of fuzzy sets, Fuzzy sets and systems* 90 (2) (1997) 141–150.
[27] L. A. Zadeh, *Fuzzy logic= computing with words*, *Fuzzy Systems*, IEEE Transactions on 4 (2) (1996) 103–111.
[28] T. P. Martin, B. Azvine, *The X-mu approach: Fuzzy quantities, fuzzy arithmetic and fuzzy association rules*, in: 2013 IEEE Symposium on Foundations of Computational Intelligence (FOCI), IEEE, 2013, pp. 24–29. doi:10.1109/FOCI.2013.6602451.
[29] D. J. Lewis, T. P. Martin, *X-mu Fuzzy Association Rule Method*, in: Proceedings of the 13th UK Workshop on Computational Intelligence (UKCI) 2013, IEEE, 2013, pp. 144–150. doi:10.1109/ UKCI.2013.6651299.

,