Online ISSN : 0975-4172
Print ISSN : 0975-4350
DOI : 10.17406/GJCST

Global Journal

of Computer Science and Technology: H

Information & Technology

Big Data Analysis

Analysis of Salary Dataset

Highlights

Telugu Text Categorization

Artificial Intelligence and Mind

Discovering Thoughts, Inventing Future

Volume 16 Issue 4 Version 1.0

Global Journal of Computer Science and Technology: H
Information & Technology

Global Journal of Computer Science and Technology: H
Information & Technology

Volume 16 Issue 4 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: *Open Association of Research Society*
Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin: 452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Global Journals Constitutional Editorial Board

~Integrated~

Dr. Xiaohong He

Professor of International Business

University of Quinnipiac

BS, Jilin Institute of Technology; MA, MS, Ph.D,

(University of Texas-Dallas)

Web: quinnipiac.edu/x1606.xml

Dr. T. David A. Forbes

Associate Professor and Range Nutritionist

Ph.D Edinburgh University - Animal Nutrition

M.S. Aberdeen University - Animal Nutrition

B.A. University of Dublin- Zoology.

Web: essm.tamu.edu/people-info/faculty/forbes-david

Dr. Burcin Becerik-Gerber

University of Southern Californi

Ph.D in Civil Engineering

DDes from Harvard University

M.S. from University of California, Berkeley

M.S. from Istanbul Technical University

Web: i-lab.usc.edu

Dr. Bassey Benjamin Esu

B.Sc. Marketing; MBA Marketing; Ph.D Marketing

Lecturer, Department of Marketing, University of Calabar

Tourism Consultant, Cross River State Tourism Development Department

Co-rdinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Söhnke M. Bartram

Department of Accounting and Finance

Lancaster University Management School

Ph.D. (WHU Koblenz)

MBA/BBA (University of Saarbrücken)

Web: lancs.ac.uk/staff/bartras1/

Dr. Maciej Gucma

Asistant Professor,

Maritime University of Szczecin Szczecin, Poland

Ph.D. Eng. Master Mariner

Web: www.mendeley.com/profiles/maciej-gucma/

Dr. Söhnke M. Bartram

Ph.D, (IT) in Faculty of Engg. & Tech.

Professor & Head,

Dept. of ISE at NMAM Institute of Technology

Dr. Shun-Chung Lee

Department of Resources Engineering,

National Cheng Kung University, Taiwan

Dr. Balasubramani R

Department of Accounting and Finance

Lancaster University Management School

Ph.D. (WHU Koblenz)

MBA/BBA (University of Saarbrücken)

Web: lancs.ac.uk/staff/bartras1/

Dr. Fotini Labropulu

Mathematics - Luther College, University of Regina

Ph.D, M.Sc. in Mathematics

B.A. (Honours) in Mathematics, University of Windsor

Web: luthercollege.edu/Default.aspx

M. Meguellati

Department of Electronics,

University of Batna, Batna 05000, Algeria

Dr. Vesna Stanković Pejnović

Ph. D. Philospohy , Zagreb, Croatia

Rusveltova, Skopje, Macedonia

Contents of the Issue

Global Journal of Computer Science and Technology: H Information & Technology
Volume 16 Issue 4 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Big Data Analysis using Spark for Collision Rate Near CalStateLA

By Manik Katyal, Parag Chhadva, Shubhra Wahi & Jongwook Woo
California State University

Abstract- Police say alcohol, drugs and speed are the three major factors that cause collisions, we thought that it would be insightful to analyze the collision data to ensure the correctness of this conclusion; and also to get further information like what age groups were involved, in what areas have accidents occurred, what were the reasons behind collisions, etc. These experiences can possibly make overall population mindful of the reasons for crashes created by impacts.

To analyze more than hundred thousand records we adopted Spark for faster processing of this massive data set. In this paper, we are presenting facts based on data and analytics which lead to conclusions like the number of collisions decreased between 2009 and 2013, Females involved in collisions were much less than males, etc. Moving ahead in our research, we addressed complex analytics like areas near CalStateLA more prone to collisions, brands of cars more involved in collisions and which specific type of collision was most observed.

Keywords: *spark, collision data, gender analysis, geo spatial analysis, big data.*

GJCST-H Classification: *C.2.1,C.2.3*

Strictly as per the compliance and regulations of:

Big Data Analysis using Spark for Collision Rate Near CalStateLA

Manik Katyal [α], Parag Chhadva [σ], Shubhra Wahi [ρ] & Jongwook Woo [ω]

Abstract- Police say alcohol, drugs and speed are the three major factors that cause collisions, we thought that it would be insightful to analyze the collision data to ensure the correctness of this conclusion; and also to get further information like what age groups were involved, in what areas have accidents occurred, what were the reasons behind collisions, etc. These experiences can possibly make overall population mindful of the reasons for crashes created by impacts.

To analyze more than hundred thousand records we adopted Spark for faster processing of this massive data set. In this paper, we are presenting facts based on data and analytics which lead to conclusions like the number of collisions decreased between 2009 and 2013, Females involved in collisions were much less than males, etc. Moving ahead in our research, we addressed complex analytics like areas near CalStateLA more prone to collisions, brands of cars more involved in collisions and which specific type of collision was most observed.

Keywords: spark, collision data, gender analysis, geo spatial analysis, big data.

I. Introduction

Big Data is defined as non-expensive frameworks that can store a large scale data and process it in parallel [9, 10].

There are huge number of collisions reported around the world and the death and serious injury rates are significantly growing as indicated by statistics and evaluations published by different agencies across the world. Collisions can happen because of many reasons like alcohol, drugs, circumstances, inattentiveness, etc. We have used DataBricks Spark over Hadoop to analyze our data because of the following reasons:

- Apache Spark promises speed up to 100 times faster than Hadoop MapReduce and comfortable APIs.
- Spark can run as a standalone or on top of Hadoop YARN, where it can read data directly from HDFS.
- Apache Spark processes data in-memory while Hadoop MapReduce persists back to the disk after a map or reduce action, so Spark should outperform Hadoop MapReduce.
- Spark has comfortable APIs for Java, Scala and Python, and also includes Spark SQL. Hadoop MapReduce, however, is written in Java and is infamous for being very difficult to program.

Author α σ ρ ω: Department of Computer Information Systems, College of Business and Economics, California State University, Los Angeles.
e-mail: swahi@calstatela.edu

According to the Statewide Integrated Traffic Records System (SWITRS) there were a total of 173,715 collisions reported through the years 2009-2013 in the LA i.e. an average 34,745 number of collisions each year and around 95 collisions each day. Our analysis is primarily focused on SWITRS data for the years 2009-2013 for Los Angeles in order to provide insight into the types of collisions, areas with higher frequency of collisions, and time windows of high frequency collisions etc. This can potentially enable law enforcement divisions in crash apprehension, collision reduction, and in turn take proactive measures to control the collision rate and ensure safety of drivers and spread awareness among drivers.

Our dataset is large and unstructured. Hence, storing data in Hadoop Distributed File System was an option. But when compared we came down to a conclusion that Hadoop is scalable, but the programming API is lacking. Real-time streaming frameworks like Storm are a good fit for fixed processing from firehoses, but not so much for flexible queries from a data store. Thus, we have turned to Spark, a fast in-memory distributed computing framework, to help us with these insights. The utility of using Spark to consume large amounts of data and enable analytical insights is what we aim to demonstrate.

II. Related Work

Cameron [8] and Andrew et al [9] also performed a details analysis of crash data in their studies. The difference in our approach is that for our research we adopt big data techniques and principles for data analysis. It is not only to store and process massive data set but also faster to analyze such a massive data set using these techniques. Both papers will be discussed in detail below in terms of technology used, data source, etc.

First, we will discuss the Accident Data analysis done by Cameron. Cameron targeted the data set by breaking it down into several clusters. In this paper, clustering technique was used to divide data into several buckets like motorcycle, trucks, cars, pedestrians, etc. This clustering technique leaded to constructive analysis of the traffic data of Victoria, Europe. The analysis was focused around counter measurements (a measure that attempts to break the road trauma chain before happen) and target groups (which specifically targeted a segment of vehicle on the

road). A brilliant way to look at the problem and some great conclusions have been derived.

In comparison to Cameron's research this paper is targeting one of the busiest cities of the America in terms of traffic, Los Angeles. Apart from this, the main thought about the approach is different. This paper discusses about the level of severity, highest type of collision, gender analysis, etc. and not dividing data into segments of type of vehicle driven. Also, we have used Spark, a fast big data analytical tool that made processing much faster and efficient. We also used better graphical representations like 3-D maps and heat maps.

Another similar research study was done by Andrew who, unlike Cameron, divided hos data set into hot spots. Hots Spots means the areas where most collisions occurred like bridges, intersections, road segments to increase police awareness in those areas. This is similar to the approach we have adopted in this paper. Andrew's paper covers Tuscaloosa County, Alabama and they developed a new Geographic Information System which pointed hot spots in various areas of the Tuscaloosa County. On the contrary, we did not build any tool for analysis and showcasing real time change in information. We discussed the various factors that cause collision.

III. Our Big Data Systems

We have utilized the SWITRS 2009-2013 collision dataset [1] for Los Angeles city to do point by point examination on crash patterns, geographic occurrence, recurrence of collision and so on.

To add an alternate measurement to the investigation we considered California State University Los Angeles and Downtown Los Angeles as our point of convergence and continued to extend the information in light of various parameters like time and distance. This would bring about extricating key discoveries about crashes happening around our university and in Los Angeles.

We stored our unstructured data in the data bricks cluster. Data Bricks provides reliable and scalable storage for storing and analyzing unstructured data and also runs on the top of the Spark. The advantage of making use of the Spark is the In-Memory computation. Data stored in Data Bricks Cloud and can be queried using different programming languages, for our analysis we used iPython Notebook to extract the data as it provides an SQL-like structure and operates on tables just like an RDBMS, even though it does not strictly follow the full SQL standard.

We have used Data Bricks for our analysis as one of the advantages provided by Data Bricks is a cluster that can be launched, dynamically scaled up or down, and terminated with just a few clicks. One other key advantage is the ability to importing data seamlessly from sources like Amazon S3, local machines, or other data sources including HDFS, RDBMS, Hive, Cassandra and Mongo DB.

- Cluster Size = 6GB
- No. of Nodes = 5
- No of CPU cores = 8
- Driver Speed = .88 cores
- Spark Version = 1.6.1

Our datasets in approximately 170 MB and consists of total 173,715 rows in comma separated values (csv) format, which provides information on type of collision, area, time, location and date and other information regarding the collision. We also had two other files for Victims and Parties affected in the collisions which have around 1.5 million records.

IV. Collision Data Analysis in Spark

It was observed that there was a decrease in total number of collisions reported from 2009 till 2013 and this can be seen in Figure 1.

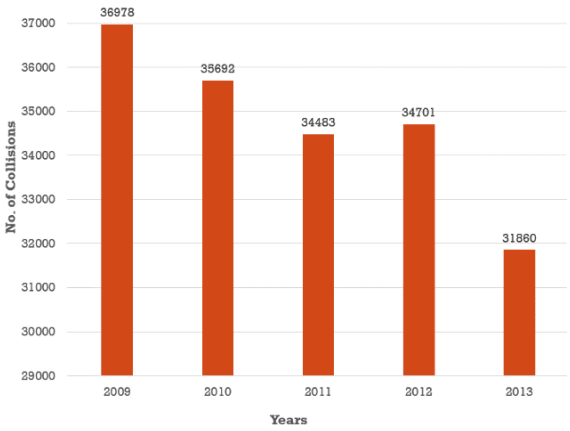


Figure 1: Collision rate decrease over the years

Further that in Figure 2 shows that from all the kind of collisions reported, the most kind of collisions which were committed were broadside collisions followed by Rear side collisions, Sideswipe and Hit object as shown in the graph below:

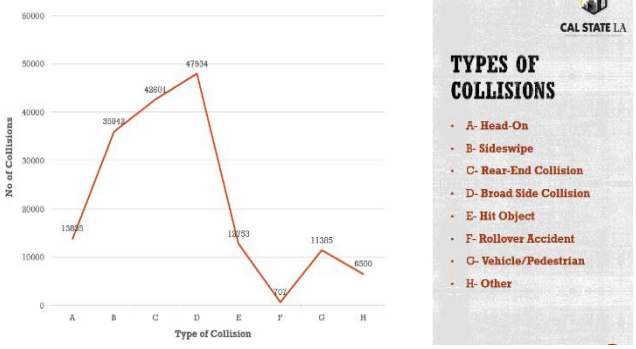


Figure 2: Types of collisions

a) Data Visualization

Data visualization is projecting the relevant information in pictorial or graphical form. iPython query results can be visualized in Microsoft Excel. We have used Power View for data visualization in excel. Apart from Excel there are other powerful tools available for data visualization/dashboard creation like R and Tableau.

b) Graphical representations of data analyzing results

The other important parameters in our analysis were day and time in Figure 3, where we analyzed collisions that occurred each day and for every two-hour interval. We observed that collisions were higher between Friday and Monday than the rest of the days.

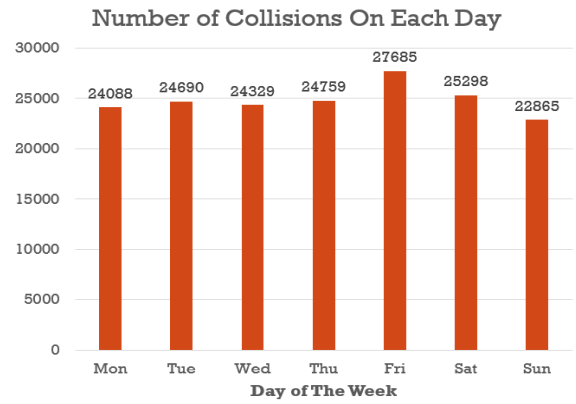


Figure 3: Day analysis of Collisions Reported

Another thing to observe was that the most number of collisions were reported during 12:00 PM to 8:00 PM where peak time of collisions is between 2PM to 6PM. The least is during 4:00 AM to 6:00 AM. This is shown in Figure 4.

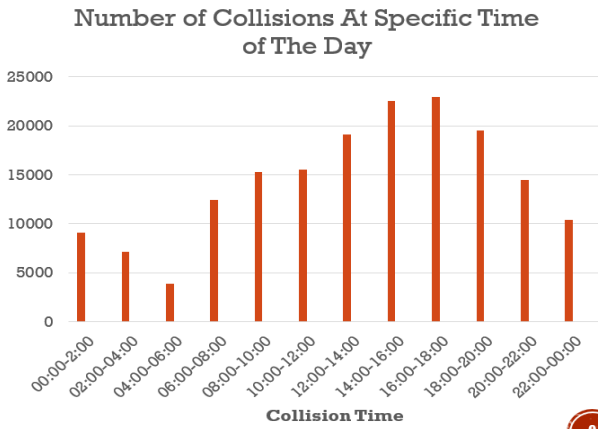


Figure 4: Collisions Reported in 2 hour intervals

We performed Gender wise analysis as well and found that 60% of the collisions that were reported had a male driver driving the car, in only 40% of the cases women drivers were driving the car as shown in Figure 5.

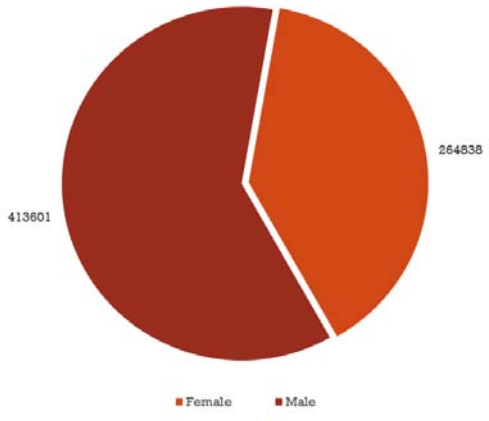


Figure 5: Gender Wise Collision Analysis

Our data set included a column for age of parties involved in collisions and we extracted this data to include another measurement for age wise analysis of collisions.

The different age groups data covered by the data set was identifies to varied and ranging between infants and senior citizens. When the entire age group data was analyzed, we found that most parties that were involved in the collisions were between the age of 20-30. Results also did not leave any uncertainty and demonstrated that youth was clearly involved in most collisions which can be seen in Figure 6.

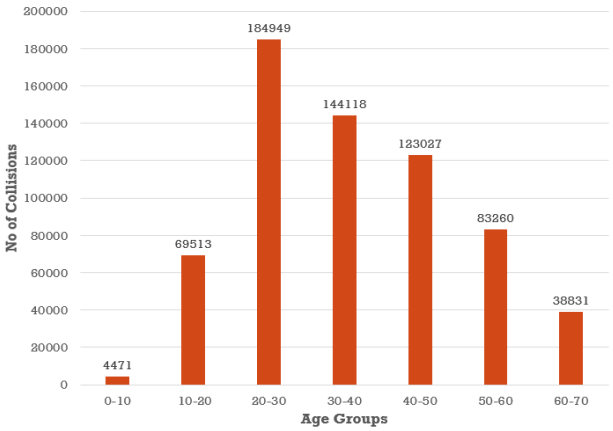


Figure 6: Age of Parties Involved in Collision

It was insightful to perform a level of severity analysis and it was interesting to find that collisions that have been reported over the years were either at level 4 of severity or at level 0 which means either fatal or mere property damage which are both extreme cases. The Figure 7: below shows this analysis.

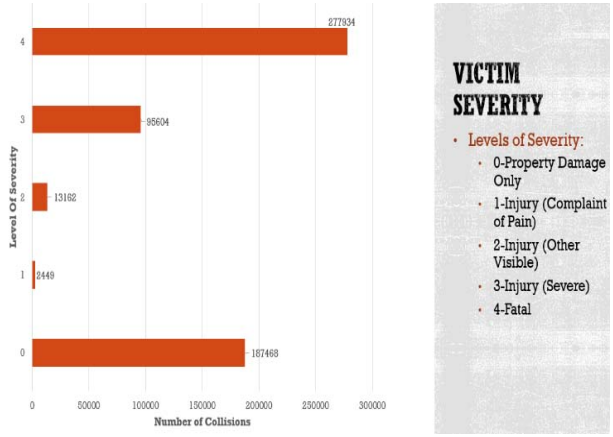


Figure 7: Victim Severity

We also did a car wise analysis to see which cars were most involved in collisions and to make it interesting we bifurcated the cars into consumer segment (which are affordable and seen on the road more) and the luxury section (which are expensive cars). It was interesting to find that Toyota being most loved as a car actually was involved in most collisions followed by Ford and Honda in the consumer segment as seen in Figure 8. Toyota has sold around 5.6 million cars between 2009 & 2013.

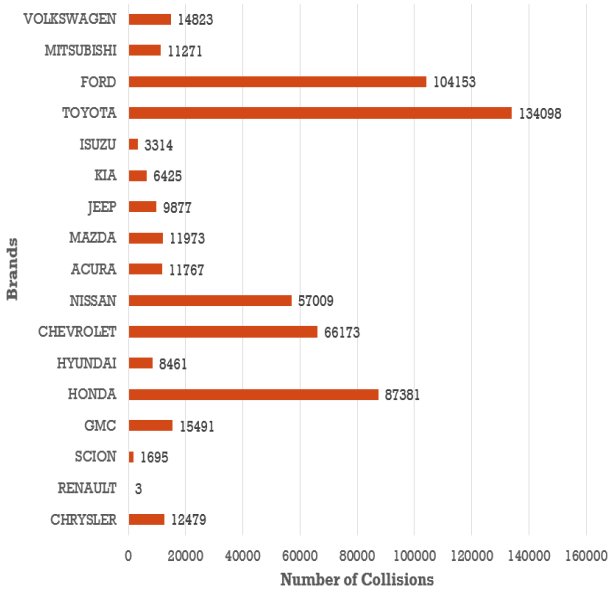


Figure 8: Toyota beats all other cars in no. of collisions

As per Figure 9, in the luxury car segment Mercedes-Benz, BMW and Cadillac was most involved in collisions.

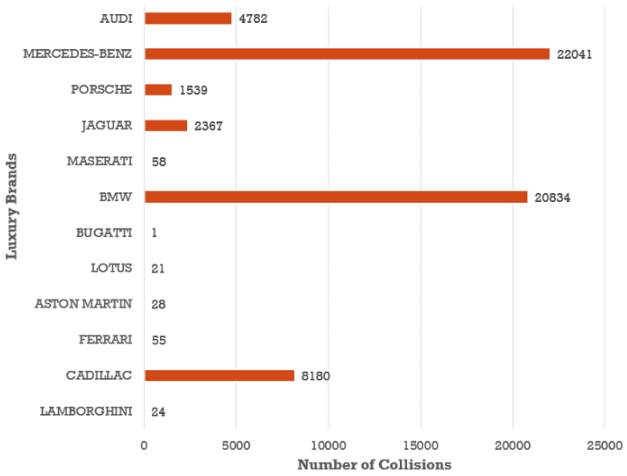


Figure 9: Collisions reported in luxury car brands

Figure 10 clearly illustrates the analysis based on seasons, it made us affirmatively say that Fall season sees most collisions and summer sees the least.

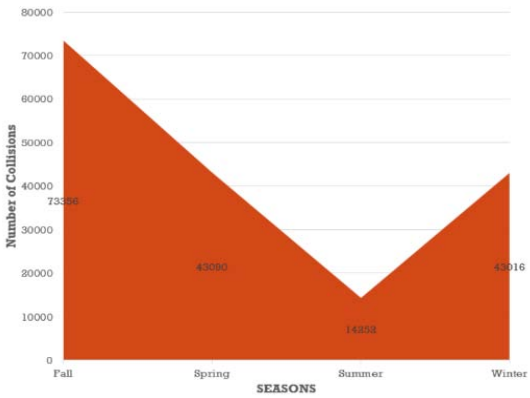


Figure 10: Collisions as reported in different seasons

Alcohol is a concern associated with driving and collisions world-wide. Based on our analysis, we see that driving under the influence of alcohol involved more severe collisions though he number of collisions under the influence of alcohol have decreased since 2009 which is a good sign but the number is still high standing at around 3265 collisions in 2013 as illustrated in Figure 11. As per our analysis Figure 11 shows that weekends were prone to having collisions involving alcohol as compare to rest of the week.

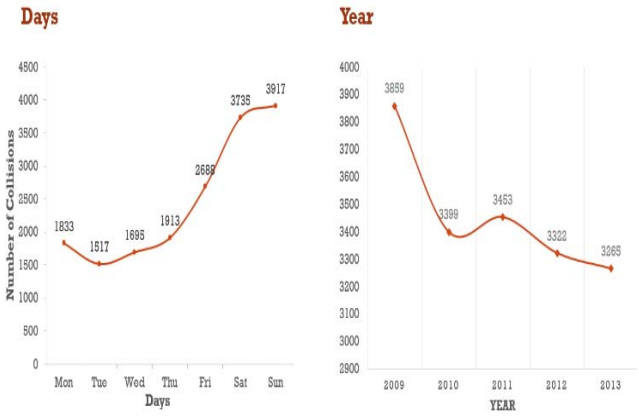


Figure 11: Collisions under the Influence of Alcohol

Figure 12 shows as per the collision rate analysis approximately 34% of the cases were Critical in terms of level of severity when alcohol was involved in the situation of collision.

Table 1: Level of Severity

Level of Severity **	Description
0	Property Damage Only
1	Injury (Complaint of Pain)
2	Injury (Other Visible)
3	Injury (Severe)
4	Fatal

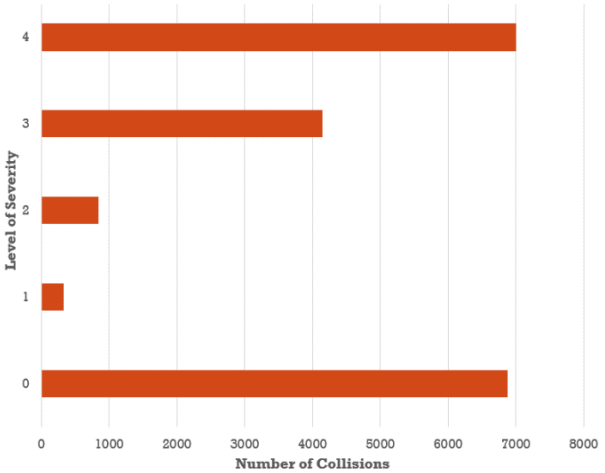


Figure 12: Collision Severity under the influence of alcohol

We took our university as a focal point and calculated the distance to every collision reported from our university, we found that the most of number of crimes were committed within the radius of 5-10 miles. Clearly, Figure 13 indicates that within 10 miles of our campus there are more collision incidents then areas far around.

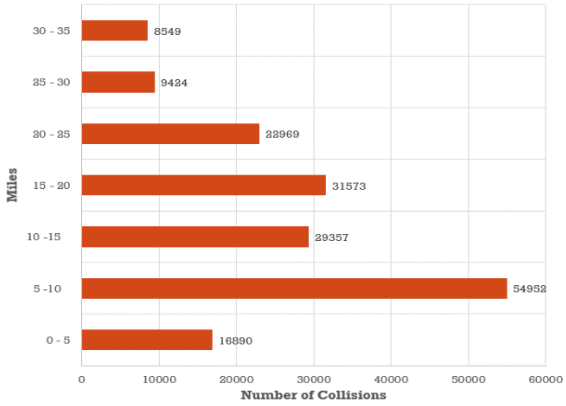


Figure 13: Collisions setting CSULA as the focal point

To make this more visual and interesting we created a heat map using excel power view which clearly shows the area around CSULA and its frequency. The Figure 14 below show red areas which mean high number of collisionsreported and the areas where color fades from yellow to green shows the frequency has decreased.

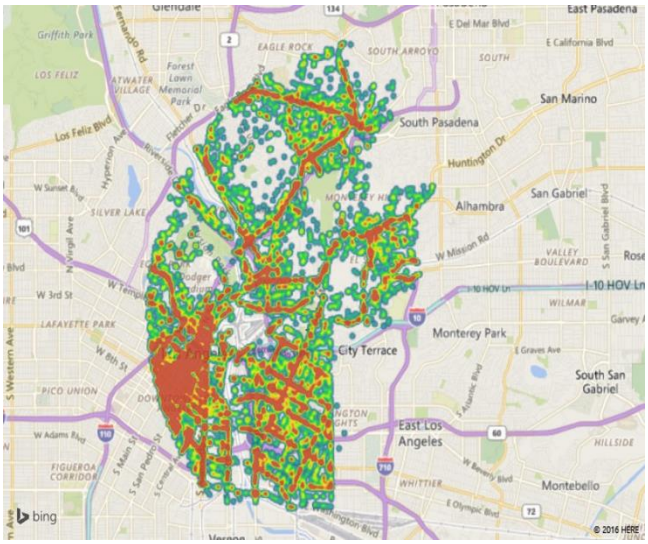


Figure 14: Heat map showing collisions in areas around CSULA

Figure 15 clearly illustrates that as per the collision data analysis, the rate of collision is more on freeways like i10, i5, 710, i405 & i210 which are closer to CSULA, Down Town LA, Pasadena, Burbank, Hollywood & Santa Monica.

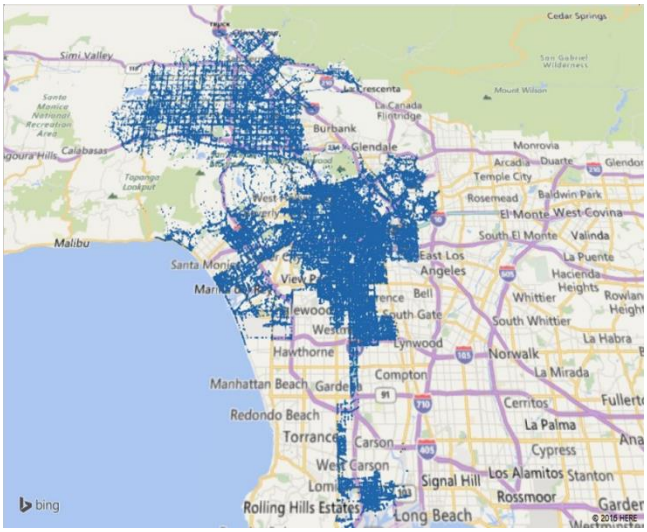


Figure 15: Geographical representation of all the Collisions reported in the LA County.

As per the analysis, Figure 16 shows more male drivers being involved in severe fatal collisions as compared to female drivers (Refer Table 1)

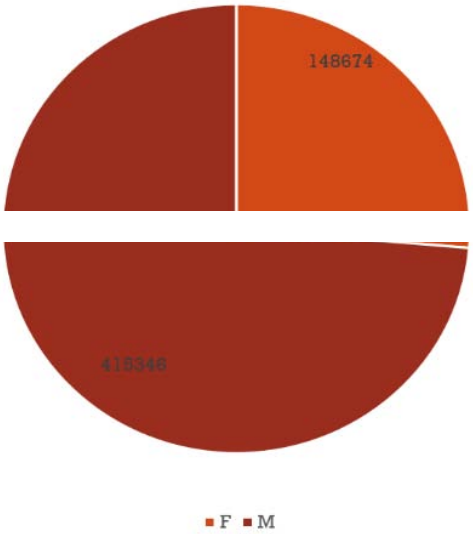


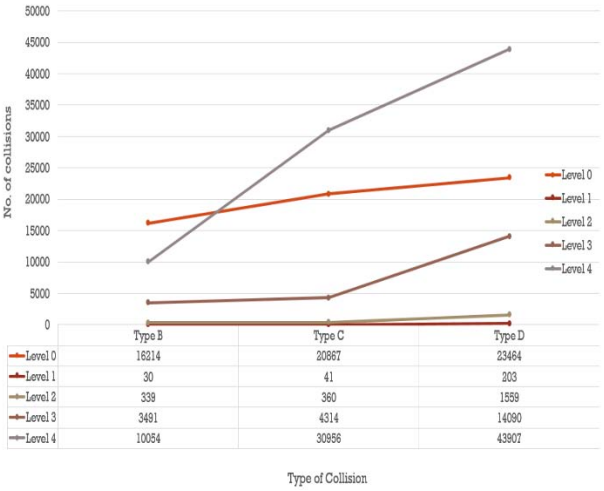
Figure 16: Gender analysis when level of severity is 4

Table 2: Type of Collision

Type of Collision	Description
A	Head-On
B	Sideswipe
C	Rear-End Collision

Type of Collision	Description
D	Broad Side Collision
E	Hit Object
F	Rollover Accident
G	Vehicle/Pedestrian
H	Other

As per our analysis shown in Figure 17 it is clear that Type B, C & D (described in Table 2) saw most collisions which were fatal too (For level of severity refer Table 1).



	Type B	Type C	Type D
Level 0	16214	20867	23464
Level 1	30	41	203
Level 2	339	360	1559
Level 3	3491	4314	14090
Level 4	10054	30956	43907

Figure 17: Levels of collision analysis in the highest type of collisions reported.

4.4 3-D Maps

Our visualization tool has the ability to make data visualization a lot more interactive. One such powerful feature of the tool is the 3D maps which present results on a multidimensional or perhaps to present geo spatial data. We decided to utilize this feature and create 3D map for the degree of severity of injury (refer Table 1) shown in Figure 18 below:

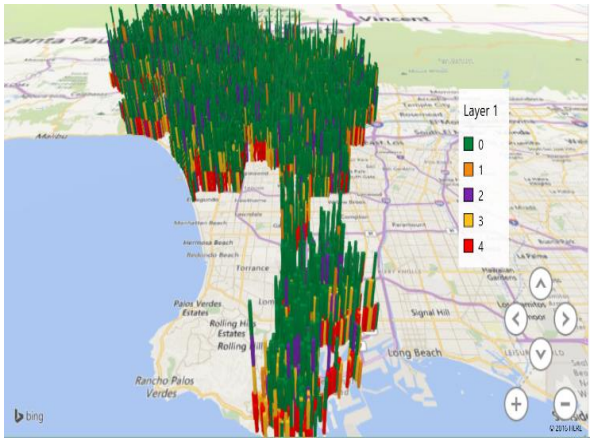


Figure 18: 3D view of collisions in LA

To make the map more readable in terms of what the maps are trying to symbolize, we decided to focus on the Long Beach area which is a coastal city of the Los Angeles county and a port in Southern California. The graph shows the different levels of severity of collisions that occurred on a geo location, Figure 19 and 20 shown below describe these:

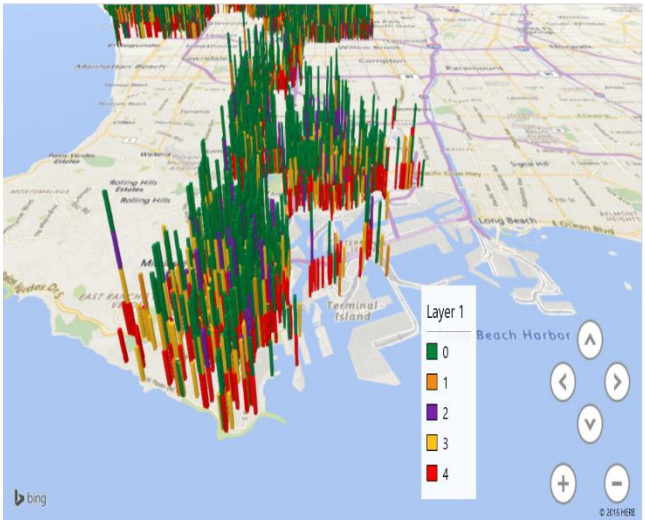


Figure 19: 3D view of collisions in Long beach area

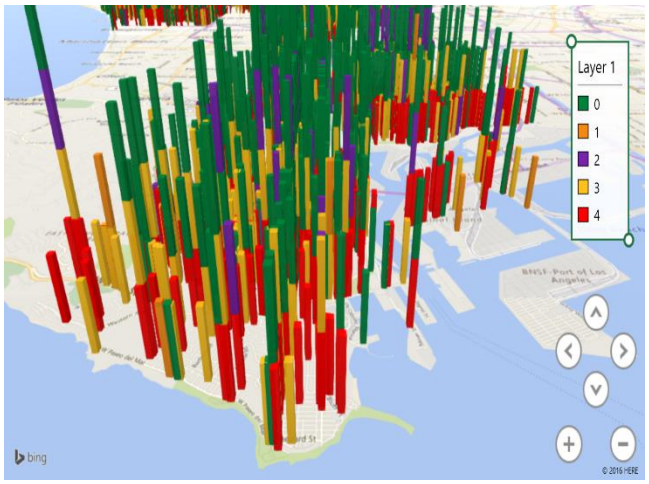


Figure 20: Zoomed in view of the above map

V. Conclusion

In this paper we have adopt the approach to leverage the Data Bricks and its Spark platforms for big data analysis. We have queried the data stored in Data Bricks cloud and using the iPython Notebook to query the data. Also we used Databricks and its benefits when compared to Hadoop which has more limitations.

In our detailed analysis, we were able to project and visualize data about collisions that occurred in Los Angeles. Through our analysis we found that:

1. Collisions have decreased each year from 2009 to 2013.
2. Collisions are higher between Friday and Monday than the rest of the days.
3. Most collisions occur between the time brackets of 12PM to 8PM where peak
4. Time of collisions is between 2PM to 6PM.
5. Broadside collisions occur more than other kind of collisions.
6. The age group of 20-30 is most involved in collisions.
7. Male drivers cause more collisions than female drivers. To be specific the ratio is 60:40.
8. 49% all collisions reported are critical and have been rated level 4 in terms of severity.
9. Collisions under the influence of alcohol has decreased in 2013 compared to 2009.
10. However, 34% of all collisions occurred under the influence of alcohol were fatal.
11. Fall season saw more collisions as compared other reason round the year.
12. Approximately there were 16k collisions within 5 miles of CSULA between 2009 & 2013.
13. More Collisions on freeways closer to CSULA.

References Références Referencias

1. SWITRS Dataset. http://geohub.lacity.org/ Datasets ?q=safety&geometry=-178.039%2C-7.751%2C-2.786%2C40.95
2. Collision GitHub Code. https://github.com/ paragch hadwa/ CIS520CollisionsinLA, you guys names
3. Official Apache Hadoop Website. https://hadoop. apache.org/
4. Apache Hive TM. https://hive.apache.org/
5. Python Programming Guide. https://spark.apache. org/docs/0.9.1/python-programming-guide.html
6. Introduction to Spark with Python. http://www.Kdnug gets.com/2015/11/introduction-spark-python.html
7. Spark SQL Programming Guide.1 https://spark. Apa che.org/docs/1.2.0/sql-programming-guide.html
8. M. Cameron, Monash, University Accident Research Centre - Reports 46 & 47 – 1992, Monash University Accident Research Centre "Display and Analysis of Crash Data", Dr. Andrew Graettinger, Dr. Jay K. Lindly and Mr. Gautam J. Mistry, UTCA Report Number 03102 April 1, 2005.
9. "Market Basket Analysis Algorithm with Map/Reduce of Cloud Computing", Jongwook Woo and Yuhang Xu, The 2011 international Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2011), Las Vegas (July 18-21, 2011).
10. "Market Basket Analysis Algorithms with Map Reduce", Jongwook Woo, DMKD-00150, Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery, Oct 28 2013, Volume 3, Issue 6, pp445-452, ISSN 1942-4795.

This page is intentionally left blank

Global Journal of Computer Science and Technology: H Information & Technology
Volume 16 Issue 4 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Telugu Text Categorization using Language Models

By Swapna Narala, B. Padmaja Rani & K. Ramakrishna
JNTU College of Engineering

Abstract- Document categorization has become an emerging technique in the field of research due to the abundance of documents available in digital form. In this paper we propose language dependent and independent models applicable to categorization of Telugu documents. India is a multilingual country; a provision is made for each of the Indian states to choose their own authorized language for communicating at the state level for legitimate purpose. The availability of constantly increasing amount of textual data of various Indian regional languages in electronic form has accelerated. Hence, the Classification of text documents based on languages is crucial. Telugu is the third most spoken language in India and one of the fifteen most spoken language n the world. It is the official language of the states of Telangana and Andhra Pradesh. A variant of k-nearest neighbors algorithm used for categorization process. The results obtained by the Comparisons of language dependent and independent models.

Keywords: *text categorization, language dependent and independent models, k-nearest neighbors.*

GJCST-H Classification: *D.2.11,D.2.12*

Strictly as per the compliance and regulations of:

Telugu Text Categorization using Language Models

Swapna Narala [α], B. Padmaja Rani [σ] & K. Ramakrishna [ρ]

Abstract- Document categorization has become an emerging technique in the field of research due to the abundance of documents available in digital form. In this paper we propose language dependent and independent models applicable to categorization of Telugu documents. India is a multilingual country; a provision is made for each of the Indian states to choose their own authorized language for communicating at the state level for legitimate purpose. The availability of constantly increasing amount of textual data of various Indian regional languages in electronic form has accelerated. Hence, the Classification of text documents based on languages is crucial. Telugu is the third most spoken language in India and one of the fifteen most spoken language n the world. It is the official language of the states of Telangana and Andhra Pradesh. A variant of k-nearest neighbors algorithm used for categorization process. The results obtained by the Comparisons of language dependent and independent models.

Keywords: text categorization, language dependent and independent models, k-nearest neighbors.

I. Introduction

Now a day's huge amount of information is being posted on to the web. In order to get useful information from the web, the information available has to be categorized. Text Categorization is the task of automatically categorizing a set of unlabeled text documents to their corresponding categories from a predefined category set [2]. These categories can be viewed as a set of documents and test document can be treated as a query to the system. The measures to evaluate the information retrieval systems are often applicable to measure effectiveness text categorization systems [1]. Text categorization has many applications [2], like information retrieval system, search engine, text filtering, word sense disambiguation, language identification, POS tagging and machine translation etc.

Telugu is one of the old and traditional languages of India and it is categorized as one of the Dravidian language family unit with its own high-class script. It is the authorized language of the Telangana and Andhra Pradesh states in south India. Amit et al [6] surveyed that in India the Telugu native speakers are above 50 million. It was positioned between13 to 17 largest spoken languages all over the world. Telugu is a rich morphological language that has high word conflation [7]. Various approaches for text categorization have been done on Indian languages. Most of the works have been reported on Telugu language. M Narayana Swamy et al have used KNN, NB and decision tree classifier [4]. They have experiment on Kannada, Tamil and Telugu corpus statistics is illustrated by Zipf's law. Analysis of N-gram model on text classification was proposed in the work of [5]. Goverdhan. A Durga k et al [3] projected a technique with ontology text categorization for Telugu digital-items and retrieval system. For the best of our knowledge, this is the first time our proposed language models have been applied for Telugu text categorization. The paper is structured as follows; section 2 describes the system overview, section 3 explains Testing and results and at the last, a section 4 conclusion is drawn.

II. System Overview

The system design of the proposed approach can be shown in the Figure.1. First read a text document from corpus and each line is pre-processed by elimination of non-Telugu characters, numerals and special characters like colons, semicolons and quotes. Then a pre-processed document is tokenized and extracts the raw words. Words in Telugu text are separated by spaces and are extracted with spaces as delimiter from the document and place all raw words in *Input File*. Language dependent and independent models are takes raw words from *Input File* as input. Read one word at a time from file. Finally find the root word by applying various models like vibhaktulu based stemming, suffix removal stemming, Rule based suffix removal stemming, N-gramming, pseudo N-gramming and Rule based Pseudo N-gramming. Finally, apply the text categorization.

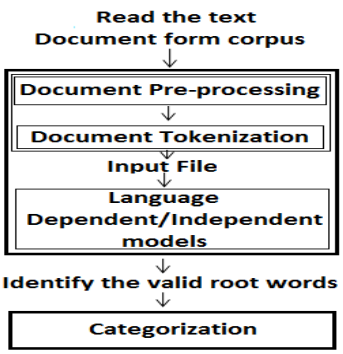


Figure 1: Proposed Approach

Author α: Research scholar, JNTU College of Engineering, Hyderabad TS, India. e-mail: swapnanaralas@gmail.com
Author σ: Department of CSE, JNTU College of Engineering, Hyderabad, TS, India. e-mail: padmaja_jntuh@yahoo.co.in
Author ρ: Department of CSE, CMR College of Engineering, Hyderabad, TS, India. e-mail: krkrishna.cse@gmail.com

a) Proposed language models

Our proposed language models are categorized in three ways are shown in figure 2. These models take raw words from *Input File* as input and identify the root word.

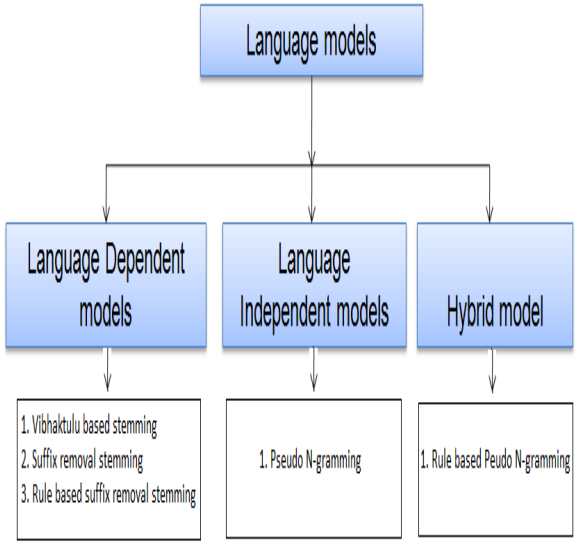


Figure 2: Proposed language models

b) Vibhaktulu based stemming

Vibhaktulu based stemming is a language dependent model. It is the process of finding the root word by removing the last one or more syllables from the word, which are matched with Telugu vibhaktulu. It is observed that, processing the complete set of input words, only 19 to 20% of words with the last syllables are matched to Telugu vibhaktulu.

c) Suffix removal stemming

Suffix removal stemming is the process of finding the root word from the word by removing the matched suffix with suffix list which is shown in figure 3. By observing the Telugu data set, it is found that maximum suffix length will be 2(two) and minimum is one. Suffix removal stemming method giving better performance than vibhaktulu based stemming algorithm. It's accuracy is 58-59%.

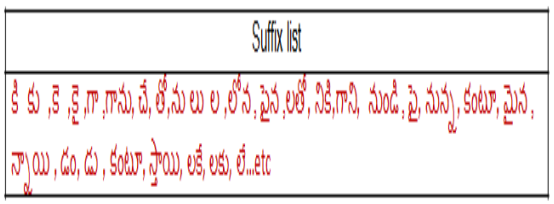


Figure 3: Suffix list

d) Rule based suffix removal Stemming

Suffix removal stemming is a base method for Rule based Suffix removal stemming algorithm. The result of suffix removal stemming words may normally contain inflections. The inflections in the stem word cannot be removed using simple suffix removal. We have designed rule based suffix removal of some possible inflections that frequently occur in the Telugu Language. The rules are used to replace characters are presented in Table 1. By these rules the electiveness of the proposed Rule based Suffix removal stemming algorithm is increased. Accuracy of Rule based suffix removal is 69-70%.

Table 1: Rules for Replacement Syllables

S.No	List of characters/syllable sound found as suffix	Replacement characters
1	అ,ఆ	అం, ఉ, ఇ
2	ఇ ,ఐం	ఉ, అ, అం
3	ఉ , ఉ + లు , ఓ +లు	ఇ, అం
4	ఎ ,ఏ, ఎం	ఇ ,ఉ ,అ, అం
5	ఒ ఓ ఔ	ఇ ,ఉ
6	అం	ఉ

e) Pseudo N-gramming

Pseudo N-gram is the process of finding the root word by stripping the word from the end. Stripping length will be taken depending on the word length. Maximum stripping length is 5 and minimum is 2. Example of Pseudo N-gramming is shown in figure 4. It is a language independent.

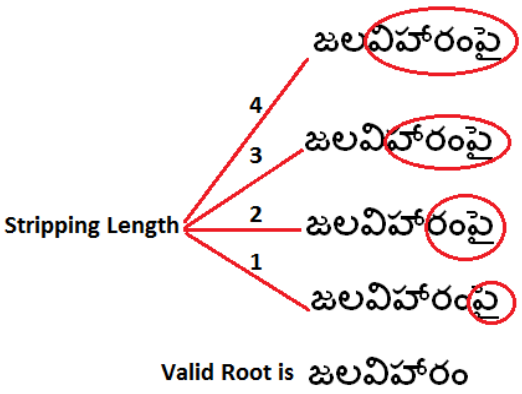


Figure 4: Pseudo N-gramming

A sequenece of words from the *Input File* was used in identifying the valid root by pseudo N-gram algorithm and the results are presented in Table 2,which contains list of words with intial & final stripping length and final valid root word.

Table 2: Result of Pseudo N-gramming

List of Words before pseudo N-gram	Intial Word length	Intial Stripping Length	Final stripping length to make a Valid Word	Stripped Suffix	Valid Root word
పాలసముద్రంలో	6	4	1	లో	పాలసముద్రం
ఏనుగులతో	5	4	2	లతో	ఏనుగు
మదపుటేనుగు	6	4	0	---	మదపుటేనుగు
భార్యలైన	4	3	2	లైన	భార్య
జలవిహారంపై	6	4	1	పై	జలవిహారం
సరోవరానికి	6	4	0	----	Not a Valid Root
నిలబడ్డాయిగాని	7	5	2	గాని	నిలబడ్డాయి
తోచలేదు	4	3	0	---	తోచలేదు
తప్పించాలో	4	3	1	లో	తప్పించా
అల్లకల్లోలం	5	4	0	---	అల్లకల్లోలం
బుద్ధిపుట్టి	4	3	2	పుట్టి	బుద్ధి
చెల్లాచెదరుగా	6	4	1	గా	చెల్లాచెదరు

f) Rule Based Pseudo N-Gramming

It is a hybrid model. Pseudo N-gram is a base method for this processing to remove suffixes from words. The result of Pseudo N-gram of some words normally contains inflections. The inflections in the stem word cannot be removed using simple Pseudo N-gram. We have designed rule based Pseudo N-gram which contain set of rules used to replace characters. These rules used for words normally contain more inflections that frequently occur in the Telugu Language. List of rules with sample example are shown in Table 3.

Table 3: List of rules for Rule based pseudo N-gramming

S.No	List of characters/syllable sound found as suffix	Replacement characters	List of Words are not recognized by Pseudo N-gram	List of words recognized by Rule based Pseudo N-gram
1	అ,ఆ	అం, ఉ, ఇ	చెప్పడానికి ప్రమాదాన్ని కెరటాల వర్షానికి గంపెడాశ పెళ్ళయిన	చెప్పడం ప్రమాదం కెరటం వర్షం గంపెడు పెళ్ళి
2	ఇ ,ఇం	ఉ, అ, అం	నిషేదించడం దేవుడి హింసించే	నిషేదం దేవుడు హింస
3	ఉ , ఉ + లు , ఓ +లు	ఇ, అం	ఆకృతుల్ని ఎడారులు సన్నజాజుల చీకట్లో	ఆకృతి ఎడారి సన్నజాజి చీకటి
4	ఎ ,ఏ, ఎం	ఇ ,ఉ ,అ, అం	చోటక్కడ మురిపెంగ పనిమిలదు ఎక్కడక్కడో	చోటు మురిపం పని ఎక్కడ
5	ఒ ఓ ఔ	ఇ ,ఉ	కాలోకటి తాడోపెడో	కాలు తాడు
6	అం	ఉ	పగలంతా కళ్ళం తా	పగలు కళ్ళు

g) K-NN Classifier

The k-NN classifier is a similarity-based learning method that has been shown to be very effective for a variety of problem domains including text categorization [9, 10]. Given a test document, the k-NN method finds the k nearest neighbors among the training documents, and uses the categories of the k neighbors to weight the category. The similarity score of each and every neighbor document to the test document is used as the weight of the classes of the neighbor document.

III. Testing and Results

The proposed models are evaluated on Telugu Corpus, collected from online newspapers and Wikipedia. This work has been implemented on sample selection of 1,500 documents of seven categories are presented in Table 4.

Table 4: Categories of Telugu Documents

	క్రీడలు	పాటలు	కథలు	సాహిత్యము	వార్తలు	రాజకీయాలు	నడులు
No. of. Doc	244	110	120	247	100	268	80
Words	44,233	10,640	88,427	26,255	87,552	99,964	27,061

To evaluating the performance of the proposed system using KNN classification, we use the typical evaluation metrics that come from information retrieval – precision (P), recall (R), and F1 measure:

$$P = TP /(TP + FP) \quad \text{.....(1)}$$

$$R = TP/(TP + FN) \quad \text{..... (2)}$$

$$F1 = (2*P*R) /(P + R) \quad \text{.....(3)}$$

Where TP is True Positives, TN is True Negatives, FN is False Negatives and FP is False Positive [8]. We have projected the performance of the proposed language models result with KNN classifier shown in Table 5.

Table 5: Performance of Language models

Language Model	Recall	Precision	F1 Measure
Vibhaktulu based stemming	75.22	86.4	81.94
Suffix removal stemming	76.42	87.5	81.58
Rule based suffix removal stemming	77.11	78.01	77.56
N-Gramming	78.02	84.30	81.03
Pseudo N-Gramming	79.96	82.77	81.34
Rule based Pseudo N-Gramming	82.81	87.77	85.22

Both Recall and Precision are to be high for an efficient performance. F1 measure reflects the overall accuracy. Recall and Precision graph is shown in figure 5(a) and 5(b). From the results we observed that, Rule based Pseudo N-gram model has high precision and recall. So it is efficient model for text categorization.

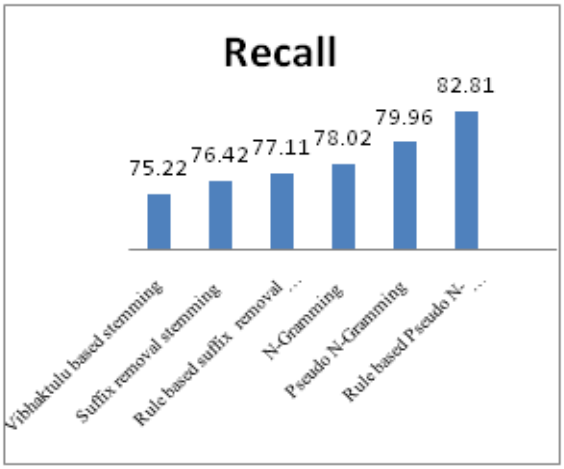


Figure 5(a): Recall Graph

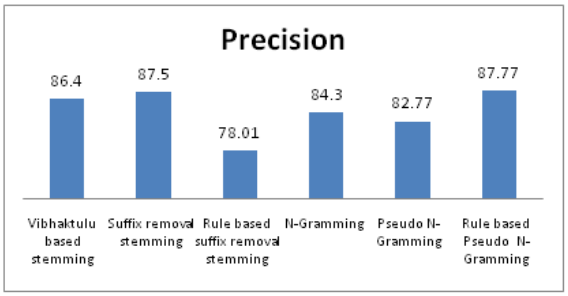


Figure 5 (b): Precision Graph

IV. Conclusion

In this paper, we proposed various language dependent and independent models. Among these models the performance of Rule based pseudo N-gramming is more. So it is well suited for Telugu Text categorization. As part of our research work in Telugu categorization, it is also suitable for other complex Indian languages like Hindi, Malayalam and Kannada.

References Références Referencias

1. T. K. Landauer, P. W. Foltz, D. Laham, "An Introduction to Latent Semantic Analysis", Discourse Processes, 1998, pp. 259-284.
2. Murthy K.N., Automatic Categorization of Telugu News Articles, Department of Computer and Information Sciences, University of Hyderabad, Hyderabad,2003.
3. Mrs. A. Kanaka Durga, Dr. A. Govardhan, Ontology Based Text Categorization Telugu Documents, International Journal of Scientific & Engineering Research Volume 2, Issue 9, September- 2011 , PP: 1-4
4. Indian Language Text Representation and Categorization Using Supervised Learning Algorithm M Narayana Swamy1 ,M. Hanumanthappa
5. Vishnu Vardhan B.Analysis of N-gram model on Telugu Document classification thesis ,2008.

6. A D Manning, P. Raghavan, and H. Scutze., An introduction to information retrieval, Cambridge: Cambridge university press, Vol. 1, 2009, PP:6.
7. U.Rao, 2008, Functional Specifications of Morphology CLATS, Hyderabad Central University, Version 1.3.1, 2008, PP:1-32.
8. Joachims, Thorsten: Learning to classify text using support vector machines: Meth- ods, theory and algorithms. Kluwer Academic Publishers, (2002).
9. Yang, Y. and Liu, X. (1999). A Re-examination of Text Categorization Methods. In Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval, pages 42-49.
10. Mitchell, T.M. (1996). Machine Learning. McGraw Hill, New York, NY.

This page is intentionally left blank

Global Journal of Computer Science and Technology: H Information & Technology
Volume 16 Issue 4 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Artificial Intelligence and Mind

By Dr. Nergis Ustoglu
Istanbul Maltepe State Hospital, Istanbul Turkey

Abstract- According to the absence of any indication in favor of the subjective experience in artificial intelligence simulations including a model with many numbers of neurons and synapses, this should be due to the system structure, not the numbers. This structure may be due to an extra dimension with a different information (quale). While the number of transistors is directly proportional to the more complex calculation, it does not include “awareness”.

GJCST-H Classification: *I.2,I.2.11*

Strictly as per the compliance and regulations of:

Artificial Intelligence and Mind

Dr. Nergis Ustoglu

Abstract- According to the absence of any indication in favor of the subjective experience in artificial intelligence simulations including a model with many numbers of neurons and synapses, this should be due to the system structure, not the numbers. This structure may be due to an extra dimension with a different information (quale). While the number of transistors is directly proportional to the more complex calculation, it does not include "awareness".

I. Introduction

Artificial intelligence researchers, tried to mimic the activities carried out by the brain in terms of the functional aspects. Alan Turing, claims that if the machine replaced by human can convince the interrogator that he is a human, then someone can conclude that the machine is thinking. Can a machine that has knowledge of all things have human awareness?

Using C2 onLLNL's Dawn Blue Gene/P supercomputer with 147.456 CPUs and 144 TB of main memory, the DARPA's neuromorphic adaptive SyNAPSE, that has the ambitious goal of engendering a revolutionary system of compact, low-power neuromorphic and synaptronic chips by using novel synapse-like nanodevices, is compared with the number of neurons and synapses in cortices of mammals [1].

Cortical simulation which exceeds that of a cat cerebral cortex is achieved. A cat with 6.10×10^{12}synapses is simulated using a model, with 0.9×10^{9} neurons and 0.9×10^{13}synapses, that uses probabilistic connectivity and a simulation time step of $1\,ms$, only 83 times slower than real time per Hertz of average neuronal firing rate. They were succesful in memory storage, uncovering the the relationship in data and in pattern recognition but could not have the consciousness of a cat with subjective experience.

If we use "Occam's Razor", is it possible to say that the awareness of a system, which is non-connected to a processor belongs to the structure of the system, even though the connection between network and data processing is clear?

One of the co-founders of phenomenology, Franz Brentano [2], speaks of a science based on internal perception. Mental acts can totally cover themselves by folding over themselves and taking themselves as an object. Mind can become conscious about a perception in the mind by taking it as an object.

In artificial intelligence, though self-folding loop continues forever; it will never be possible. Neither the multiplicity of synapses nor the excess number of processors can be the solution. As expressed in "Gödel's incompleteness theorem", the proof of "inconsistency" in a system will not be possible within the system itself. We all perceive three spatial and one time dimensional space, thus, within this system, it is not possible to understand the "awareness". So our perception, our memory and self-experience of our concepts need a more advanced system with an additional dimension.

The algorithms used in artificial intelligence are formed up on dimensions that we can only perceive and which are non-overlapping with the subjective experience. Unlike a computer, the brain internalization of an apple in the memory could be just the case supplied by an additional spatial dimension, without depending on a function. The fourth dimension, which can be visualized as a matter which is intertwined in three dimensions, can be interpreted as the "awareness", like the formation of matter in three dimensions.

The missing side of artificial neural networks, processing at nano seconds, compared to human brain at milliseconds, may be some structural properties in the hardware, not the computation.

Artificial neural networks are mathematical systems in a neuromorphic network of process units which are connected in a weighted manner. Despite Moore's law and the huge development in computational analysis and the memory storage, nothing has been achieved in terms of awareness and self-sufficiency. Similarly, a more enlarged space is required for the algorithms which are more complex and sensitive; like the homunculus in a brain.

In a living thing with nervous system, unlike a computer, all of these algorithms work with input data, each with a certain "awareness". To feel, to suffer, to get pleasure, to know what you think, require an extra information (awareness). This information (quale), which is so different than what we perceive, could be due to some extra data given by neurons on a micro level. It is not possible to observe the "awareness" within the system, by these brain-made algorithms in the 3 spatial and 1 time dimensions. While quale is being formed in that extra dimension, plasticity and computations are done in the sub-dimension.

Consciousness is a case which develops its algorithm with the connections between the neurons after taking the inputs, which can turn back into and cover themselves, supplied by the extra dimension of

Author: *Istanbul Maltepe State Hospital, Altaycesme Mah. Cam Sok. No:26-28 34843 Maltepe / İstanbul-TURKEY December 2016.*
e-mail: vunal@yeditepe.edu.tr

different laws. Sensation, awareness, that are experienced in this extra dimension, would also supply the autonomy of mind.

II. Conclusion

The fact that artificial intelligence simulations have no sense and awareness is not about the number of transistors and the network but about the extra dimension where "quale" is being formed.

References Références Referencias

1. "The cat is Out of the Bag: Cortical Simulations with 10^9 Neurons, 10^{13} Synapses", Rajagopal Anantha-narayanan, Steven K. Esser, Horst D. Simon, and Dharmendra S. Modha, SC09 November 14-20, 2009, USA
2. Brentano, F. (1874) "The distinction between Mental and Physical Phenomena", Philosophy of Mind, Classical and Contemporary Reading, Davis Chalmerz, Oxford University Press, 2002.

Global Journal of Computer Science and Technology: H Information & Technology
Volume 16 Issue 4 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Big Data Analysis of Salary Dataset using Hive

By Ishan Fafadia
California State University Los Angeles

Abstract- One way to understand how a city government works is by looking at who it employs and how its employees are compensated. This data contains the names, job title, and compensation for San Francisco city employees on an annual basis from 2011 to 2014. The analyzed data will be shown in the form of various charts and graphs with respect to 1. Yearly Mean Pay, 2. Mean Pay by Job Type, 3. Pay based on Base Pay, Overtime Pay, Other Pay and Benefits. As the Salary seeking population grows, the data also grows in size. This becomes a challenge for the traditional RDBMS to manage the huge volumes of data. Hence Salary data Analysis can be made using Hive and Map Reduce algorithms to eliminate the challenges faced by the traditional RDBMS.

GJCST-H Classification: C.2.1,C.2.3

Strictly as per the compliance and regulations of:

Big Data Analysis of Salary Dataset using Hive

Ishan Fafadia

Abstract- One way to understand how a city government works is by looking at who it employs and how its employees are compensated. This data contains the names, job title, and compensation for San Francisco city employees on an annual basis from 2011 to 2014. The analyzed data will be shown in the form of various charts and graphs with respect to 1. Yearly Mean Pay, 2. Mean Pay by Job Type, 3. Pay based on Base Pay, Overtime Pay, Other Pay and Benefits. As the Salary seeking population grows, the data also grows in size. This becomes a challenge for the traditional RDBMS to manage the huge volumes of data. Hence Salary data Analysis can be made using Hive and Map Reduce algorithms to eliminate the challenges faced by the traditional RDBMS.

- We have observed the drop of budget allocation of salaries in San Francisco.
- There were some departments which didn't provide any benefits to their employers.
- For some departments, even if the employer had worked overtime they were not paid for their extra work.
- Good thing that we observed is that there was no gender discrimination among the department.

I. Introduction

A standout amongst the most well-known datasets urban areas ordinarily discharge is their compensation structure.

Thus we grabbed the dataset of San Francisco as it is the most essential city for any graduate understudy. What's more, we discovered some conceivably intriguing edges of investigation:

1. How has pay rates changed after sometime between various Departments of individuals?
2. How are base pay, extra minutes pay, and advantages apportioned between various gatherings?
3. Is there any proof of pay separation taking into account sexual orientation in this dataset?
4. How spending plan is distributed in light of various Department and obligations?
5. And In this project we have focused on the payment structure of the considerable number of divisions and attempt to give the answer for low paying office.

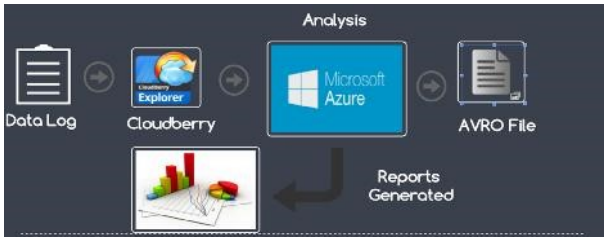
Hadoop is an open source, Java-based programming structure that backings the handling and capacity of to a great degree substantial information sets in a disseminated figuring environment. Hadoop makes it conceivable to run applications on frameworks with a huge number of product equipment hubs, and to handle a large number of terabytes of information.

Author: e-mail: ifafadi@calstatela.edu

Apache Hive is an information distribution center framework based on top of Hadoop for giving information synopsis, query, and analysis. Hive gives a SQL-like interface to inquiry information put away in different databases and document frameworks that incorporate with Hadoop.

II. Work Flow

Initially a data set with Employee_Id, EmployeeName, JobTitle, BasePay, OvertimePay, Other Pay,Benefits, TotalPay, TotalPayBenefits, Year, Notes, Agency,Status is taken from an authentic source. As a next step, this comma separated file has to be uploaded to the cloud. This is done with the help of cloud berry explorer. And data is converted to Avro format.



a) Data Storage

We changed over our information in Avro Format and we utilize that same information we put away in cloudberry explorer and we are utilizing avro in light of the fact that, Avro is one of the favored information serialization frameworks in view of its language lack of bias.

Because of absence of language versatility in Hadoop writable classes, Avro turns into a characteristic decision as a result of its capacity to handle various information designs which can be further prepared by different languages.

b) Conversion to Avro Format

To change over csv information to Avro information utilizing Hive we have to take after the progressions beneath:

1. Make a Hive table put away as content document and indicate your csv delimiter too.
2. Load csv document to above table utilizing "load data" command.
3. Make another Hive table utilizing AvroSerDe.
4. Embed information from previous table to new Avro Hive table utilizing "insert overwrite" command.

c) Data Representation

In this Project, we have considered four main parameters to Analyze the data.

1. Change of Maximum Total Pay in 4 Years.
2. Change of Mean Pay Yearly
3. Mean Pay of each Department
4. Benefits given in each Year.
5. Payment Structure of each Department.

These data were obtained by writing suitable queries in Hive QL.

1. Select Year, max (TotalPay) from avro_table group by Year order by 1
2. Select year, percentile (cast (Totalpay as bigint), 0.5), count (*) Records from avro_table group by year order by 1;
3. Select Job Type, percentile (cast (Totalpay as bigint), 0.5), count (*) as Records from avro_table; JobType group by JobType;
4. Select Year, Sum (Benefits) from avro_table group by year order by 1
5. select JobType, cast(avg(Basepay) as bigint), cast (avg(Overtimepay) as bigint),cast(avg(Otherpay) as bigint), cast(avg(Benefits) as bigint) from avro_table JobType group by JobType;

III. Data Analysis

a) Change of Maximum Total Pay in 4 Years

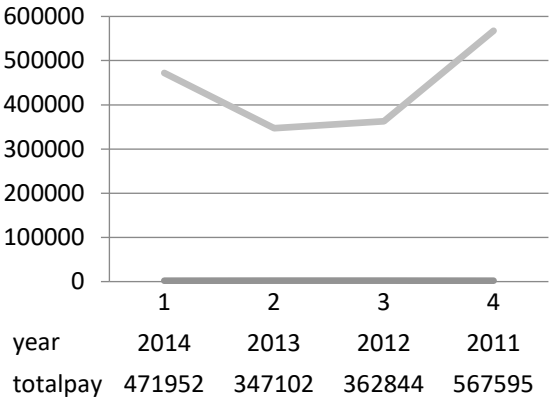


Figure 1

From Figure 1, it's clearly shown that there was a drastic changes of Maximum Total pay from the year 2011 to 2012 which continues till 2013 but it was risen in the year 2014 .From this we can inferred that in year 2012 and 2013 there was slight recession and because of that payment structure of employees were not increase.

b) Change of Mean Pay Yearly

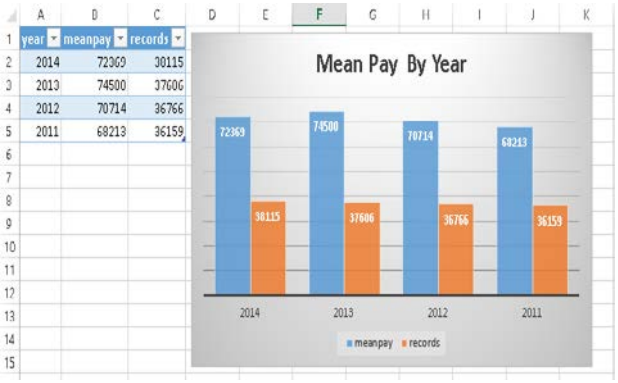


Figure 2

From Figure 2 we come to know that the mean pay was increased leaving the fact that maximum total pay was decreased and this could be possible because the whole budget was well distributed among the employees and number of employees was also increased so San Francisco hired more employees from 2011 to 2014 so because of that more people were employed and benefited And there was drop of budget allocation from 2011 to 2014.

c) Mean Pay of each Department

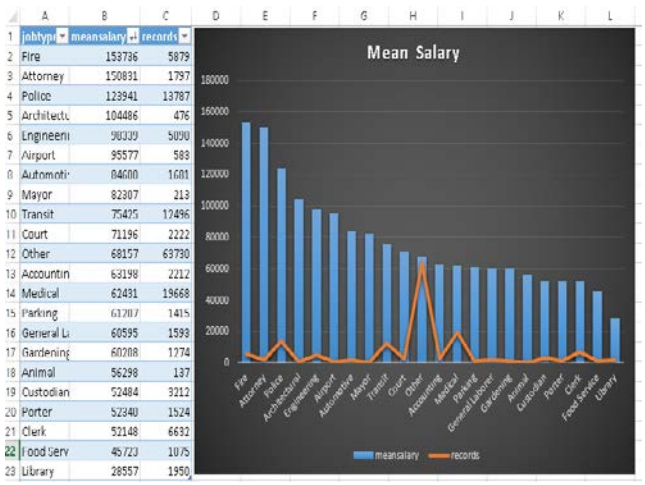


Figure 3

In Figure 3, we have shown Job Type, Mean pay & records.

And to find the best paying department we need to find the Mean pay for each department and the number of records in it.So here we found that fire department has the highest Mean pay.

And we know that in any city Fire department is the most important group of professionals as they serve for day and night at any situation and library department is the least paying department. And one thing we found surprising that medical department is also not a good pay master and because of this less people are interested in taking medical as their career

d) Benefits given in each Year

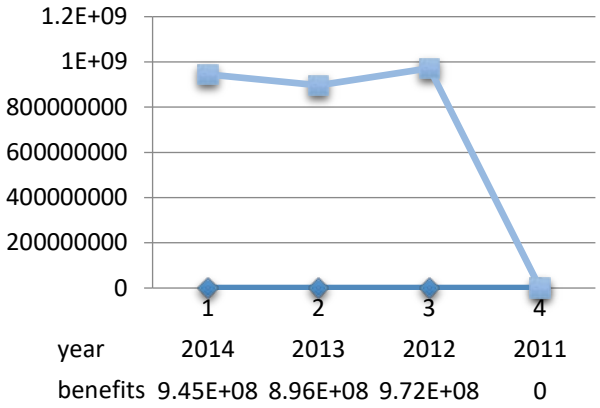


Figure 4

In Figure 4, we try to find the benefits that was given to employees in all the year and we find that in the year 2011 no benefits were given by any department and this could be reason of least mean pay in the that year

But by the year 2012 there was the added pay in the name of benefits to the payment structure of each department to lure more employees and provide better living standard to the people.

Payment Structure of each Department

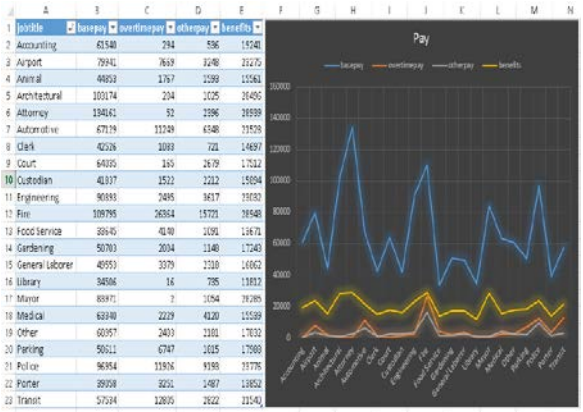


Figure 5

Figure 5 shows the basepay, overtimepay, otherpay & benefits.

From this we come to know that Attorney has got the highest base pay and Food Service got the lowest. But there was not so much difference in Overtime,Otherpay and Benefits.

IV. Business Solution

All the departments should have better structure for overtime pay & other pays.

We can reallocate the budget to the smaller departments.

We can reduce the payment structure of Attorney, Police and Fire.

And to encourage employees we should provide the equality among all people regarding their post and stature Example:

As we all the Engineering department is also very nowadays because of growth of computers in every field so to increase the mean pay of that department we can increase the overtime pay and other pay which can make them in top 4 earning department of San Francisco.

And medical department is deep low in the mean pay so to uphill their department we can increase the base pay which is very low in terms of their dedication and risk in their works and by doing that we can encourage more people to join the medical line in future.

For some departments, even if the employer had worked overtime they were not paid for their extra work. And from this we come to know that main reason of fire department having best mean pay is there they have the best structure of Overtime pay, Benefits, Other pay.

V. Conclusion

We have observed the drop of budget allocation of salaries in San Francisco.

There were some departments which didn't provide any benefits to their employers.

For some departments, even if the employer had worked overtime they were not paid for their extra work.

Good thing that we observed is that there was no gender discrimination among the department.

And Fire Department is the best in the San Francisco area.

VI. Future Work

With the dataset we had, we analyzed the salaries based on different departments. But if we had bigger data i.e. if we had data of last 15-20 years we would have more precisely provided results about departmental salaries.

And with more precise data we could have shown some better solution for the employees working in their respective departments and for the departments as well And seeing the future prospect of our analysis we can say that San Francisco government can use this to decide all the future payment structure of all a departments to provide better life and better living standards for people And seeing the future prospect of our analysis we can say that San Francisco government can use this to decide all the future payment structure of all a departments to provide better life and better living standards for people

Github Code:
https://github.com/saket18/sfsalariesanalysis

Dataset URL:
https://www.kaggle.com/kaggle/sf-salaries

References Références Referencias

1. https://hadoop.apache.org/ https://hadoop.apache.org/
2. https://hive.apache.org/
3. http://hortonworks.com/hadoop-tutorial/how-to-processdata-with-apache-hive/
4. http://blog.cloudera.com/wp-content/uploads/2010/01/6IntroToHive.pdf
5. http://blog.cloudera.com/blog/2013/04/demoanalyzing-data-with-hue-and-hive/
6. https://azure.microsoft.com/enus/documentation/articles/hdinsight-hadoop-tutorial-getstarted-windows/
7. https://azure.microsoft.com/enus/documentation/articles/hdinsight-connect-excel-hiveodbc-driver/
8. https://azure.microsoft.com/en-us/documentation/articles/hdinsight-connect-excel-powerquery/
9. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial. html
10. http://www.infoworld.com/article/2683729/hadoop/10ways-to-query-hadoop-with-sql.html

Global Journals Inc. (US) Guidelines Handbook 2016

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

Global Journals Incorporate (USA) is accredited by Open Association of Research Society (OARS), U.S.A and in turn, awards "FARSC" title to individuals. The 'FARSC' title is accorded to a selected professional after the approval of the Editor-in-Chief/Editorial Board Members/Dean.

- The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

FARSC accrediting is an honor. It authenticates your research activities. After recognition as FARSC, you can add 'FARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, and Visiting Card etc.

The following benefits can be availed by you only for next three years from the date of certification:

FARSC designated members are entitled to avail a 40% discount while publishing their research papers (of a single author) with Global Journals Incorporation (USA), if the same is accepted by Editorial Board/Peer Reviewers. If you are a main author or co-author in case of multiple authors, you will be entitled to avail discount of 10%.

Once FARSC title is accorded, the Fellow is authorized to organize a symposium/seminar/conference on behalf of Global Journal Incorporation (USA).The Fellow can also participate in conference/seminar/symposium organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent.

You may join as member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. In addition, it is also desirable that you should organize seminar/symposium/conference at least once.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The FARSC can go through standards of OARS. You can also play vital role if you have any suggestions so that proper amendment can take place to improve the same for the benefit of entire research community.

As FARSC, you will be given a renowned, secure and free professional email address with 100 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.

The FARSC will be eligible for a free application of standardization of their researches. Standardization of research will be subject to acceptability within stipulated norms as the next step after publishing in a journal. We shall depute a team of specialized research professionals who will render their services for elevating your researches to next higher level, which is worldwide open standardization.

The FARSC member can apply for grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A. Once you are designated as FARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria. After certification of all your credentials by OARS, they will be published on your Fellow Profile link on website https://associationofresearch.org which will be helpful to upgrade the dignity.

The FARSC members can avail the benefits of free research podcasting in Global Research Radio with their research documents. After publishing the work, (including published elsewhere worldwide with proper authorization) you can upload your research paper with your recorded voice or you can utilize chargeable services of our professional RJs to record your paper in their voice on request.

The FARSC member also entitled to get the benefits of free research podcasting of their research documents through video clips. We can also streamline your conference videos and display your slides/ online slides and online research video clips at reasonable charges, on request.

The FARSC is eligible to earn from sales proceeds of his/her researches/reference/review Books or literature, while publishing with Global Journals. The FARSC can decide whether he/she would like to publish his/her research in a closed manner. In this case, whenever readers purchase that individual research paper for reading, maximum 60% of its profit earned as royalty by Global Journals, will be credited to his/her bank account. The entire entitled amount will be credited to his/her bank account exceeding limit of minimum fixed balance. There is no minimum time limit for collection. The FARSC member can decide its price and we can help in making the right decision.

The FARSC member is eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get remuneration of 15% of author fees, taken from the author of a respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

The ' MARSC ' title is accorded to a selected professional after the approval of the Editor-in-Chief / Editorial Board Members/Dean.
The “MARSC” is a dignified ornament which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., MARSC or William Walldroff, M.S., MARSC.

MARSC accrediting is an honor. It authenticates your research activities. After becoming MARSC, you can add 'MARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, Visiting Card and Name Plate etc.

The following benefitscan be availed by you only for next three years from the date of certification.

MARSC designated members are entitled to avail a 25% discount while publishing their research papers (of a single author) in Global Journals Inc., if the same is accepted by our Editorial Board and Peer Reviewers. If you are a main author or co-author of a group of authors, you will get discount of 10%.

As MARSC, you will be given a renowned, secure and free professional email address with 30 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The MARSC member can apply for approval, grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A.

Once you are designated as MARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria.

It is mandatory to read all terms and conditions carefully.

AUXILIARY MEMBERSHIPS

Institutional Fellow of Open Association of Research Society (USA)-OARS (USA)

Global Journals Incorporation (USA) is accredited by Open Association of Research Society, U.S.A (OARS) and in turn, affiliates research institutions as "Institutional Fellow of Open Association of Research Society" (IFOARS).

The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

The IFOARS institution is entitled to form a Board comprised of one Chairperson and three to five board members preferably from different streams. The Board will be recognized as "Institutional Board of Open Association of Research Society"-(IBOARS).

The Institute will be entitled to following benefits:

The IBOARS can initially review research papers of their institute and recommend them to publish with respective journal of Global Journals. It can also review the papers of other institutions after obtaining our consent. The second review will be done by peer reviewer of Global Journals Incorporation (USA) The Board is at liberty to appoint a peer reviewer with the approval of chairperson after consulting us.

The author fees of such paper may be waived off up to 40%.

The Global Journals Incorporation (USA) at its discretion can also refer double blind peer reviewed paper at their end to the board for the verification and to get recommendation for final stage of acceptance of publication.

The IBOARS can organize symposium/seminar/conference in their country on behalf of Global Journals Incorporation (USA)-OARS (USA). The terms and conditions can be discussed separately.

The Board can also play vital role by exploring and giving valuable suggestions regarding the Standards of "Open Association of Research Society, U.S.A (OARS)" so that proper amendment can take place for the benefit of entire research community. We shall provide details of particular standard only on receipt of request from the Board.

The board members can also join us as Individual Fellow with 40% discount on total fees applicable to Individual Fellow. They will be entitled to avail all the benefits as declared. Please visit Individual Fellow-sub menu of GlobalJournals.org to have more relevant details.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

After nomination of your institution as "Institutional Fellow" and constantly functioning successfully for one year, we can consider giving recognition to your institute to function as Regional/Zonal office on our behalf.
The board can also take up the additional allied activities for betterment after our consultation.

The following entitlements are applicable to individual Fellows:

Open Association of Research Society, U.S.A (OARS) By-laws states that an individual Fellow may use the designations as applicable, or the corresponding initials. The Credentials of individual Fellow and Associate designations signify that the individual has gained knowledge of the fundamental concepts. One is magnanimous and proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice.

Open Association of Research Society (US)/ Global Journals Incorporation (USA), as described in Corporate Statements, are educational, research publishing and professional membership organizations. Achieving our individual Fellow or Associate status is based mainly on meeting stated educational research requirements.

Disbursement of 40% Royalty earned through Global Journals : Researcher = 50%, Peer Reviewer = 37.50%, Institution = 12.50% E.g. Out of 40%, the 20% benefit should be passed on to researcher, 15 % benefit towards remuneration should be given to a reviewer and remaining 5% is to be retained by the institution.

We shall provide print version of 12 issues of any three journals [as per your requirement] out of our 38 journals worth $ 2376 USD.

Other:

The individual Fellow and Associate designations accredited by Open Association of Research Society (US) credentials signify guarantees following achievements:

- The professional accredited with Fellow honor, is entitled to various benefits viz. name, fame, honor, regular flow of income, secured bright future, social status etc.

- In addition to above, if one is single author, then entitled to 40% discount on publishing research paper and can get 10%discount if one is co-author or main author among group of authors.
- The Fellow can organize symposium/seminar/conference on behalf of Global Journals Incorporation (USA) and he/she can also attend the same organized by other institutes on behalf of Global Journals.
- The Fellow can become member of Editorial Board Member after completing 3yrs.
- The Fellow can earn 60% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.
- Fellow can also join as paid peer reviewer and earn 15% remuneration of author charges and can also get an opportunity to join as member of the Editorial Board of Global Journals Incorporation (USA)
- • This individual has learned the basic methods of applying those concepts and techniques to common challenging situations. This individual has further demonstrated an in–depth understanding of the application of suitable techniques to a particular area of research practice.

Note :

"

- In future, if the board feels the necessity to change any board member, the same can be done with the consent of the chairperson along with anyone board member without our approval.
- In case, the chairperson needs to be replaced then consent of 2/3rd board members are required and they are also required to jointly pass the resolution copy of which should be sent to us. In such case, it will be compulsory to obtain our approval before replacement.
- In case of "Difference of Opinion [if any]" among the Board members, our decision will be final and binding to everyone.

"

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications.

Research letters: The letters are small and concise comments on previously published matters.

5. STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

You must strictly follow above Author Guidelines before submitting your paper or else we will not at all be responsible for any corrections in future in any of the way.

Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

TECHNIQUES FOR WRITING A GOOD QUALITY RESEARCH PAPER:

1. Choosing the topic: In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. Pick a good study spot: To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. Think technically: Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page
- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.

- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.
- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

Index

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350