

## Customer Loan Prediction Analysis

Dr.SK Mahaboob Basha<sup>1</sup>, B.laxmi prasanna<sup>2</sup>,G Pavan sai<sup>3</sup>,R.Naga sahithi<sup>4</sup>, S.Priyanka<sup>5</sup>

<sup>1,2,3,4,5</sup>Department of Computer Science and Engineering

<sup>1,2,3,4,5</sup>Sree Dattha Institute of Engineering and Science, Sheriguda, Telangana.

### ABSTRACT

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this paper we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this paper is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing.

**Index Terms:** Fruit detection, automated robots, real-time processing, multi-task cascaded convolutional network, image fusion, chloroplast, chromoplast, color transition, carotenoids.

### 1. INTRODUCTION

The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician and biologist Ronald Fisher in his 1936 paper "The use of multiple measurements in taxonomic problems" as an example of linear discriminate analysis. It is sometimes called Anderson's Iris data set because Edgar Anderson collected the data to quantify the morphologic variation of Iris flowers of three related species. Two of the three species were collected in the Gaspé Peninsula "all from the same pasture, and picked on the same day and measured at the same time by the same person with the same apparatus".

The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other. The use of this data set in cluster analysis however is not common, since the data set only contains two clusters with rather

obvious separation. One of the clusters contains *Iris setosa*, while the other cluster contains both *Iris virginica* and *Iris versicolor* and is not separable without the species information Fisher used. This makes the data set a good example to explain the difference between supervised and unsupervised techniques in data mining: Fisher's linear discriminant model can only be obtained when the object species are known: class labels and clusters are not necessarily the same.

Nevertheless, all three species of *Iris* are separable in the projection on the nonlinear branching principal component. The data set is approximated by the closest tree with some penalty for the excessive number of nodes, bending and stretching. Then the so-called "metro map" is constructed. The data points are projected into the closest node. For each node the pie diagram of the projected points is prepared.

The area of the pie is proportional to the number of the projected points. It is clear from the diagram (left) that the absolute majority of the samples of the different *Iris* species belong to the different nodes. Only a small fraction of *Iris-virginica* is mixed with *Iris-versicolor* (the mixed blue-green nodes in the diagram). Therefore, the three species of *Iris* (*Iris setosa*, *Iris virginica* and *Iris versicolor*) are separable by the unsupervising procedures of nonlinear principal component analysis. To discriminate them, it is sufficient just to select the corresponding nodes on the

principal tree.

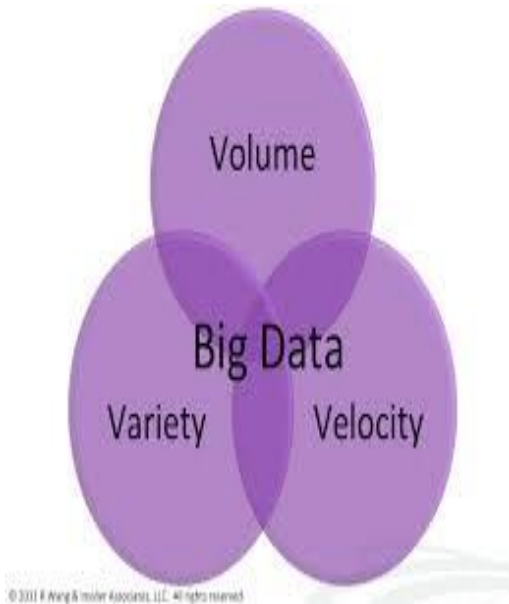
## 2. LITERATURE SURVEY

### 2.1 BIG DATA

Big data is a propelling term that depicts any voluminous measure of sorted out, semi-composed and unstructured data that can be burrowed for information. Though huge data doesn't suggest a specific sum, the term is much of the time used when discussing Petabytes and Exabyte's of data.

Bigdata is a term for informational collections that are so extensive or complex that customary information handling application programming is lacking to manage them. Gigantic data is used to depict a tremendous volume of data that is expansive to the point that it's difficult to process. The data is excessively colossal that outperforms current getting ready cutoff. Gigantic Data is an articulation used to mean an enormous volume of both sorted out and unstructured data that is so sweeping it is difficult to process using traditional database and programming frameworks. In most undertaking situations the volume of information is too enormous or it moves too quick or it surpasses current handling limit. Huge Data can possibly enable organizations to enhance operations and make speedier, more keen

choices. This information, when caught, organized, controlled, put away, and investigated can enable an organization to increase helpful understanding to expand incomes, to get or hold clients, and enhance operations.



**Fig 2.1 Big Data 3V's**

Big data can be portrayed by 3Vs: the outrageous volume of information, the wide assortment of sorts of information and the speed at which the information must be must procedures

**Volume**

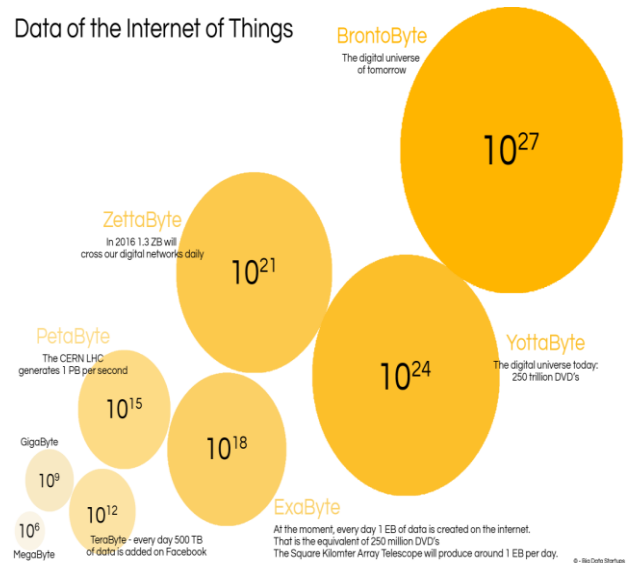
Volume is the V most connected with huge information since, well, volume can be huge. Affiliations assemble data from an arrangement of sources, including business trades, web based systems administration and information from sensor or machine-to-machine data. Beforehand, securing it would've been an issue For example, facebook stores pictures of around 250 billions.

**Velocity**

Speed is the measure of how quick the information is coming in. Information streams in at an extraordinary speed and should be managed in an opportune way. For instance, Facebook needs to deal with a torrent of photos consistently. It needs to ingest everything, process it, document it, and by one means or another, later, have the capacity to recover it.

**Variety**

Data arrives in an extensive variety of associations – from sorted out, numeric data in standard databases to unstructured substance chronicles, email, video, sound, stock ticker data and cash related trades.



**Fig 2.2 Data Measurements**

An instance of colossal data might be petabytes (1,024 terabytes) or Exabyte's (1,024 petabytes) of data involving billions to trillions of records.

### 3. SYSTEM ANALYSIS

#### 3.1 Existing System:

Machine Learning implementation is a very complex part in terms of Data analytics. Working on the data which deals with prediction and making the code to predict the future of out comes from the customer is challenging part.

#### Disadvantages of Existing System:

- Complexity in analyzing the data.
- Prediction is challenging task working in the model
- Coding is complex maintaining multiple methods.
- Libraries support was not that much familiar.

#### 3.2 Proposed System:

Python has a is a good area for data analytical which helps us in analyzing the data with better models in data science. The libraries in python makes the predication for loan data and results with multiple terms considering all properties of the customer in terms of predicting.

#### Advantages:

- Libraries helps to analyse the data.
- Statistical and prediction is very easy comparing to existing technologies.
- Results will be accurate compared to other methodologies.

### 4. SYSTEM ARCHITECTURE

#### 4.1 System Architecture:

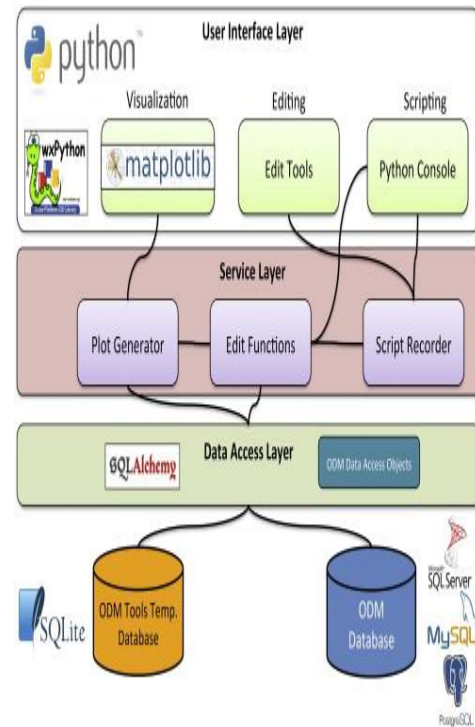


Figure 4.1 System Architecture

#### 4.2 Modules

- Extract Data
- Process the Data
- Results Generate

### 5. CONCLUSION

In this study, a new method for recognizing sentiment in iris has been proposed. The analysis shows that overall sentiment (both in iris and text) is governed by little sentiment bearing terms. In order to exploit this fact, a new method that uses Keyword Spotting (KWS) to search for sentiment bearing terms in iris has been proposed. By focusing on the terms that impact decision

and ignoring non-sentiment bearing words/phrases, the overall system is more immune to speech recognition errors. Additionally, a new method to create the sentiment bearing keyword list for KWS has also been proposed.

## 6.FUTURE ENHANCEMENT

Two of the three species were gathered in the Gaspé Peninsula "all from a similar field, and singled out that day and estimated in the meantime by a similar individual with a similar mechanical assembly.

## 7.REFERENCES

- [1] S. Johnson, "Internet changes everything: Revolutionizing public participation and access to government information through the Internet", *Administrative Law Review*, Vol. 50, No. 2 (Spring 1998) , pp. 277-337
- [2] D. Chrysanthos. "Strategic manipulation of internet opinion forums: Implications for consumers and firms." *Management Science* 52.10 (2006): 1577-1593.
- [3] M. Wollmer, et al. "Youtube movie reviews: Sentiment analysis in an audio-visual context." *Intelligent Systems*, IEEE (2013): pages 46-53.
- [4] J. Naughton, "The internet: is it changing the way we think?", *The Gaurdian*, Saturday 14 August 2010
- [5] G. Mishne and N. S. Glance, "Predicting movie sales from blogger sentiment," in *AAAI 2006 Spring Symposium on Computational Approaches to Analyzing ResMilitaris*, vol.13,n°, 3 ISSN: 2265-6294 Spring (2023)
- Weblogs, 2006.
- [6] L. Barbosa, and J. Feng, "Robust sentiment detection on twitter from biased and noisy data.", in *Proceedings of the International Conference on Computational Linguistics (COLING-2010)*. 2010.
- [7] E. Cambria, N. Howard, Y. Xia, and T. S. Chua, "Computational Intelligence for Big Social Data Analysis", *IEEE Computational Intelligence Magazine*, 11(3), 8-9, 2016.
- [8] E. Cambria, B. Schuller, Y. Xia, and B. White, "New avenues in knowledge bases for natural language processing", *Knowledge-Based Systems*, 108(C), 1-4, 2016.
- [9] M. Bautin, L. Vijayarenu, and S. Skiena. "International sentiment analysis for news and blogs.", in *Proceedings of the International AAI Conference on Weblogs and Social Media (ICWSM-2008)*. 2008.
- [10] I. Becker and V. Aharonson. "Last but definitely not least: on the role of the last sentence in automatic polarity-classification.", in *Proceedings of the ACL 2010 Conference Short Papers*. 2010.