

Bank marketing analysis using big data technology

[Business](#), [Marketing](#)



Abstract— The banking system is the backbone of every economy which is involved in lending and borrowing money when it comes to expanding of banking marketing must be required to get more customers to deposit money and the bank uses the majority of this money to lend to other customers for a variety of loans, data is getting bigger day by day and more complex when it comes with multiple attributes with traditional tools or manually it is quite impossible this problem can be solved by big data technology it is reliable to process complex data parallel in multiple machines, Bank can use this technology to constantly monitor their clients and their transactions and they can use this analysis for marketing, In this project we used MapReduce and Hive frameworks to get analysis of customers with this analysis bank can predict customers who are more likely to take loan or which client they should focus more **Keywords—**Big Data, Hadoop, MapReduce, Hive, Banking, Marketing

Introduction to provide the best knowledge to the bank from the customers' data in order to increase their revenue and quality of service we have used MapReduce framework over the customer data and Hive query to get the output as per requirement and that can be summarized or visualize in any way in this project we use Tableau and Excel Visualization tools to determine the output of our query. this document covers motivation, relevance, implementation, methods, and results this project will also highlight the future work and chances of implementation in banking marketing A.

Motivation and Relevance today most financial and banking sector working towards the data-driven approach to enhance their services and grow in the

business like other industries analytics will be a game changer in the banking sector also, as the volume of data increases it affects the level of services ...[1] with the help of big data processing techniques this can handle better, it can track customer information and their behavior simultaneously and marketing of any industry is the key factor so we have used marketing data of bank to get some analysis using the Apache Hadoop framework banks are currently shifting their marketing process to more analytical and data-driven, this analysis requires data and it is generated using the customer record and the basic marketing concepts like communication with them and campaign attempt their present financial scenario this kind of traditional methods uses by the bank and other financial institutions ...[2] if this method is applies using big data technologies so they can get more decision-making results like other marketing companies using nowadays few challenges are still there many of them working on this to get better results in this paper we have associate banking core marketing factor with technology to get analysis using big data technology B. Requirements: Big Data enables industries to customize operations and productivity this technology is capable to track real-time data and derive the best information also this technology proficient in parallel processing which facilitates to derive multiple analysis...[3] this project consists of a bank customer data with multiple attributes like occupation, age, marital status their present financial situation, currently having a loan and a few data related to marketing attempts did by the bank like response and telephonic call between customer and their marketing representatives and so on this data

can be analyzed better and many knowledge driven question we can build and classify them for the marketing perspectives.

Related Work as per Alexandra in [4]... volume, variety, and velocity these are the three core component of big data and by the time of development two more components added that is variability and value with these five dimension big data became cost-effective and decision making when it is required for predictive analysis solutions it can also integrate with advanced machine learning solutions, Combination of these two technologies proficient to provide best solutions and ground level marketing understanding. big data technology embed data into business process workflow, optimization, use cases, simulations and use case... [3] Nowadays many techniques are using for marketing strategy and most common and popular is machine learning classification and prediction their algorithm is purely based on statistics and can predict human behavior analysis, in ... [5] Hany A. Elsalamony described few data mining techniques which are best suitable for bank marketing and analysis like naive bays, decision tree model and neural networks and so on these data mining methods are very useful to get the best knowledge from customer data. Many solutions are available to make strategies for marketing solutions in machine learning and in prediction analysis machine learning algorithms are the best but it has few limitations like it take time to build the model and also data limitations are also there it can be apply for small data solutions, although banking data is much more bigger it is very difficult and expensive to get marketing solution of banking data through

machine learning so we used Big Data methods in this project to derive marketing solutions on a sample data and it can be applicable for whole data

Methodology This project is done by following few methods starting from objective, motivation and methods, methods we have done in few steps are following: A. Dataset Description: This dataset is related with direct marketing campaign of a Portuguese banking institution where customer id is removed, and no personal information is used in this dataset. Dataset is downloaded from UCI machine learning repository... [7] and it is open source data which contains details of 45212 customers with 17 attributes where 7 attributes are numeric, and 10 factors are also there with 6 categorical values and 3 binary value in yes or no and the result is also there in the form of Yes or No, detailed description of the dataset is following: Column Name Description

Column Name	Description
1. Age	Age of the customer (numeric value)
Job	Occupation of the customer (divided into 12 category)

1. Marital Marital status of the person
2. Education Qualification degree
3. Default Has credit in Default (yes or no)
4. Balance How much balance he has in account can be both positive and negative
5. Housing Customer has house or not
6. Loan Currently are they having any loan
7. Contact Their contact number if number is not known so the data is unknown

8. Month Last contact month of the year with customer
11. Day_of_week
Last contact day of the week (numeric value)
9. Duration Contact duration in seconds
10. Campaign Number of contact attempts performed to contact with customer
11. Pdays Number of days passed from the last day contact with person
12. Previous Number of contact attempts performed before the campaign
13. Poutcome Outcome of previous marketing attempts

17. y Has the client subscribed for the loan or not

Above dataset consists data of customer of a marketing campaign which contains numerical, categorical, binary and target value also this dataset is openly available for the machine learning algorithms, but we have use this because of variability of the dataset and we can apply Hadoop tools on it to derive analysis. B. Data Processing: when the data was taken from the source it was a tab-separated format we convert it to comma separated format using excel functions after that we have used some R language codes to check missing values on it we have used R because of its environment, that allow to build data cleaning scripts for data from a wide range of errors and inconsistencies...[6], but there are no missing values on that dataset that was a cleaned dataset after that we have imported the dataset into the SQL Database and saved it into SQL tables then it was added into Hadoop Distributed File System (HDFS) using Sqoop after that data imported into MapReduce function for processing or Hive database and it can transmit

afterwards into any format. C. Technology Used: Start from the data processing we have used Excel and RStudio and rest analysis part is done in the Hadoop framework using Hadoop distributed file system, Hadoop YARN, Sqoop, MapReduce and Hive MapReduce part is used in java language and IDE Eclipse used for this part that facilitates to transfer output data, and we can apply query and analysis requirements using hive language and store data into HDFS for the graphical representation using Tableau/Excel. we have used three Big Data tools MapReduce, Hive and Sqoop that are described below - i. MapReduce: MapReduce is a programming model to generate and process large datasets implemented by Google. in this framework there are two functions performed basically mapper and reducer using coding, where map process key/value pair to produce a set of intermediate key/value and another function mapper consolidate all intermediate values connected with the same intermediate key ... [8] this process is divided into four parts that are the following:

- Splitting
- Shuffling
- Mapping
- Reducing

First data splits into values and keys then the mapper function splits the data count and in terms of value later these data grouped together and generates the final output using reducer function ii. Sqoop: Sqoop tool used to transfer data from Hadoop and relational databases it is a command line interface which is used to export and import data... [9] the major ability of Sqoop is to

transfer large amount of data from Hadoop framework to database in this project Sqoop is used to import and export from the HDFS. iii. Hive Hive is the database software tool build on interface level of Apache Hadoop it can provide data summarization-based analysis, Hive interface is like SQL interface which is capable to integrate with Hadoop system. Hadoop MapReduce function can also give output based on the programmed, but it requires developers to write custom code for every output ... [10] where Hive allows users to build custom queries on data. Fig1: Hive Architecture [11] Above image explains the architecture of Hive in Hadoop system In this project Hive and SQL both have been used, but to handle the output Hive is used because it can handle large amount of data this approach is well explained in [12]

Results For this project we have performed total four analysis two are using MapReduce and two are using Hive query are following:

Task 1- the first task is to find the number of people who don't have any house and do not have any loan and when the bank representative called them for marketing purpose where the conversation between them is above 300 seconds, this task performed on MapReduce using java code in Eclipse IDE and the output file used in Excel for the column attribute assignment and visualization. For analysis we have taken number of people count with percentage and their marital status.

Task1 Above figure showing the number of people which should be the most priorily target by the bank to focus for the loan where we can see married

person in this list are the highest one so bank should make them on the top priority.

Task 2- The second task is to find the number of attempts took to convince people for the loan by the bank customer service representative grouped by their occupation this task performed through Hive query where we use job, campaign count and their outcome result sum by the number and grouped by their profession output of the Hive query is moved to HDFS where we use that file and convert it into excel format and derive graphical representation using that data.

Task 2 giving the output for the attempts where we can see management profession takes more attempts to convince for the loan, so bank have to more strategy and offer for them after management technician and admin are following top category who takes more efforts. In this category least attempts types are entrepreneur and housemaid those are two different categories but for the loan reasons differs for them like entrepreneur people requires for the developing their business, so we can say with help of this analysis bank can use different strategies as per the occupation of person.

Task 3- this task stating the outcome of the campaign on the clients where persons' age is above 50 years old and they don't have any house and loan yet although they responded well in the marketing attempts this output is grouped by the job category and represented in a chart where we can see the number of people, for this task performed using Hive query by using campaign count attempts, Housing, Loan and outcome of the customer and

their profession output of this query giving only two values job category and count of people.

Task 3 The above chart showing the number of people where we can see retired people whose age is above 50 years with no house and loan are more likely to responded in the marketing campaign After them we can see management people comes in this analysis here we can see a different kind of analysis about them in the task two they come first for taking attempts to convince for the loan and after 50 years of age they are more likely to take loan and became customer.

Task 4- Which job categories are most likely to take loan after marketing campaign this analysis is derived from the whole dataset of the customer in the MapReduce using Java in eclipse IDE only design pattern and summarization program is used to get this analysis and output saved in HDFS then after visualized using Tableau for this analysis job category and factor of campaign count is used if count is zero it means no attempts did by the bank for them so this analysis showing the people who are attended by the marketing guys and likely to taking loan.

Task 4 Task 4 diagrams stating count of people by the size of circle and color by the different category here we can see that management people are again top of the list followed by the retired, admin, blue-collar and technician. V. Conclusion the results of the bank data analysis fulfill the marketing requirements all the analysis is helpful to target customer for the loan and bank can use this analysis data for the whole data of the customer

and Big Data technology is the best choice to handle huge amount of data and helps to make strategies, from this analysis we can conclude that married person without the loan is most likely to take loans and rest analysis determining that management job profession should be the priority of the bank with the help of tasks conclusions, we can say that the marketing team should make strategies to reduce the attempts and supervise clients by their profession and previous attempts of marketing

Future Work Although Big Data technology is adapting by mostly banking and other financial intuitions and there is a lot more scope still there bank can design systems with the previous data but data is getting bigger and everyone need fast results like using real time data they can make more user friendly and increase customer service for this they can take help using social media platform, Ronald Van Loon in [13] described few solutions related to future banking methods and marketing using various social media API and real time data streaming