Employee Promotion Analysis using Python

Perikala Rajasree

Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract— *Employee sentiment analysis or Opinion Mining or Emotion Artificial Intelligence is an on-going field which refers to the use of Natural Language Processing, analysis of text and is utilized to extract quantify and is used to study the emotional states from a given piece of information or text data set. It is an area that continues to be currently in progress in field of text mining. Sentiment analysis is utilized in many corporations for review of products, comments from social media and from a small amount of it is utilized to check whether or not the text is positive, negative or neutral. Throughout this research work we wish to adopt rule- based approaches which defines a set of rules and inputs like Classic Natural Language Processing techniques, stemming, tokenization, a region of speech tagging and parsing of machine learning for sentiment analysis which is going to be implemented by most advanced python language.*

- *Human capital is of a high concern for companies' management where their most interest is in hiring the highly qualified personnel which are expected to perform highly as well. Human Resources Management (HRM) has become one of the essential interests of managers and decision makers in almost all types of businesses to adopt plans for correctly discovering highly qualified employees. Accordingly, managements become interested about the performance of these employees. Results show that professional skill development programs are needed in order to prepare employees to perform their tasks more efficiently.*
- *This Project concentrates on collecting data about employees, generating a decision tree from the historical data, testing the decision tree with attributes of an employee and generating the output as whether to give the promotion or not . The information about an employee is collected by using the user interface. This information is compared with the trained data stored in the decision tree. The final goal node is to determine whether the employee will get yearly increment, promotion or not.*

I. INTRODUCTION

A performance evaluation is a constructive process to acknowledge an employee's performance. Goals and objectives are the critical components of effective performance evaluation. The evaluation form needs to have a set of measurable goals and objectives spelled out for each area. In educational institutions, it is called as a self appraisal form. The main attributes present in the form are reading material, percentage of results, short term course attended and so on. These parameters are used to quantify the information about an employee. It acts as knowledge for taking conclusion. This paper concentrates on performance review of an employee in an educational institution. The performance evaluation is an important part in the principle of management which includes various tools for improving the performance of an employee.

1.1 Artificial Intelligence:

Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". As machines become increasingly capable, mental facilities once thought to require intelligence are removed from the definition. AI is an area of computer sciences that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include: Face recognition, Learning, Planning, Decision making etc.,

Artificial intelligence is the use of computer science programming to imitate human thought and action by analysing data and surroundings, solving or anticipating problems and learning or self-teaching to adapt to a variety of tasks.

1.2 Machine Learning

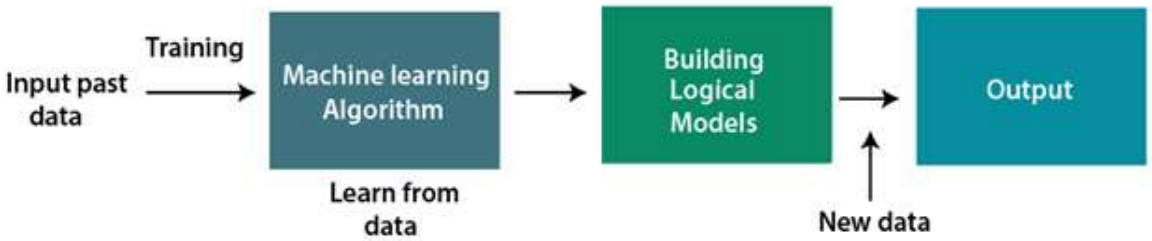
Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more.

Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. The term machine learning was first

introduced by Arthur Samuel in 1959. We can define it in a summarized way as: "Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed".

A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output. Machine learning has changed our way of thinking about the problem. The below block diagram explains the working of Machine Learning algorithm:



1.2.1 Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

1.2.2 Classification of Machine Learning

At a broad level, machine learning can be classified into three types:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

1) Supervised Learning

Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.

The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.

The goal of supervised learning is to map input data with the output data. The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher. The example of supervised learning is **spam filtering**.

Supervised learning can be grouped further in two categories of algorithms:

- **Classification**

- **Regression**

2) Unsupervised Learning

Unsupervised learning is a learning method in which a machine learns without any supervision. The training is provided to the machine with the set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data without any supervision. The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.

In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data.

It can be further classifieds into two categories of algorithms:

- **Clustering**
- **Association**

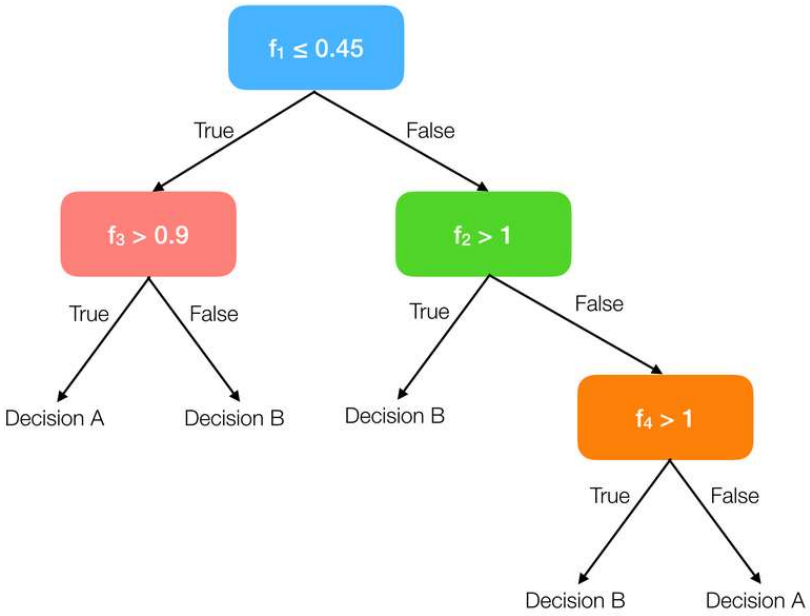
1.3 Decision Tree Classification

Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

A decision tree is a formalism for expressing such mappings. A tree is either a leaf node labeled with a class or a structure consisting of a test node linked to two or more subtrees. A test node computes some outcome based on the attribute values of an instance, where each possible outcome is associated with one of the subtrees. An instance is classified by starting at the root node of the tree. If this node is a test, the outcome for the instance is determined and the process continues using the appropriate subtree. When a leaf is eventually encountered, its label gives the predicted class of the instance. A decision tree can be constructed from a set of instances by a divide-and conquer strategy. If all the instances belong to the same class, the tree is a leaf with that class as label. Otherwise, a test is chosen that has different outcomes for at least two of the instances, which are partitioned according to this outcome. The tree has as its root a node specifying the test and, for each outcome in turn, the corresponding subtree is obtained by applying the same procedure to the subset of instances with that outcome.

In tasks with more than two classes, an alternative to growing a single tree is to construct a tree for each class that distinguishes it from all others. The idea can be taken further, encoding classes as bit strings with error correction and producing a separate tree for each bit. Higher predictive accuracy can usually be obtained by generating multiple trees from the data, all of which are used in classifying a new instance. More than one test can be used to partition the instances at each stage, giving families of superimposed trees, or multiple training sets can be samples from the data. The predictions from several trees can be combined by simple voting or by more sophisticated techniques such as stacking.

Even though the divide-and-conquer algorithm is fast, efficiency can become important in tasks with hundreds of thousands of instances or where many trees are to be produced. The most time consuming facet is sorting the instances on a numeric attribute to find the best threshold t.



II. LITERATURE SURVEY

[1] Title: Prediction of Employee Turnover in Organizations using Machine Learning Algorithms

Authors: Pankaj Ajit

Description: Employee turnover has been identified as a key issue for organizations because of its adverse impact on work place productivity and long term growth strategies. To solve this problem, organizations use machine learning techniques to predict employee turnover. Accurate predictions enable organizations to take action for retention or succession planning of employees. However, the data for this modeling problem comes from HR Information Systems (HRIS); these are typically under-funded compared to the Information Systems of other domains in the organization which are directly related to its priorities. This leads to the prevalence of noise in the data that renders predictive models prone to over-fitting and hence inaccurate. This is the key challenge that is the focus of this paper, and one that has not been addressed historically. The novel contribution of this paper is to explore the application of Extreme Gradient Boosting (XGBoost) technique which is more robust because of its regularization formulation. Data from the HRIS of a global retailer is used to compare XGBoost against six historically used supervised classifiers and demonstrate its significantly higher accuracy for predicting employee turnover**.**

[2] Title: Predicting Employee Attrition using XGBoost Machine Learning Approach

Authors: Rachna Jain; Anand Nayyar

Description: Considering the global competitive scenario, there is ocean of opportunities for skilled and talented persons in the world, and given a good chance, employees part from one organization to another. Employee turnover is regarded as the key issue for all organizations these days, because of its adverse effects on workplace productivity, and accomplishing organizational objectives on time. To overcome this problem, organizations are now taking support via machine learning techniques to predict the employee turnover. With high precision in prediction, organizations can take necessary actions at due course of time for retention or succession of employees. Most of the data comes from basic HR based database systems, which are not highly efficient in prediction and modeling and these models are not very accurate in data models and cannot assist the organizations to take successful decisions. The primary objective of this research paper is to predict employee attrition i.e. whether the employee is planning to leave or continue to work within the organization. In this paper, we propose a novel model for predicting Employee Attrition using Machine Learning based approach i.e. XGBoost which is highly robust. In order to validate the accuracy of the system proposed for Employee Attrition, the data set is acquired via online database and fetched to the system and highly stunning and precision results are shown by the system with regard to Employee turnover behavior.

[3] Title: Evaluation of machine learning models for employee churn prediction

Authors: Dilip Singh Sisodia; Somdutta Vishwakarma; Abinash Pujahari

Description: Employees are the valuable assets of any organization. But if they quit jobs unexpectedly, it may incur huge cost to any organization. Because new hiring will consume not only money and time but also the freshly hired employees take time to make the respective organization profitable. Hence in this paper we try to build a model which will predict employee churn rate based on HR analytics dataset obtained from Kaggle website. To show the relation between attributes, the correlation matrix and heatmap is generated. In the experimental part, the histogram is generated, which shows the contrast between left employees vs. salary, department, satisfaction level, etc. For prediction purpose, we use five different machine learning algorithms such as linear support vector machine, C 5.0 Decision Tree classifier, Random Forest, k-nearest neighbor and Naïve Bayes classifier. This paper proposes the reasons which optimize the employee attrition in any organization.

[4] Title: A Data-driven Analysis of Employee Promotion: The Role of the Position of Organization

Authors: Jiamin Liu; Tao Wang; Jiting Li; Jingbo Huang;

Description: In the era of big data and industry 4.0, the mode of human resource management (HRM) will be changed. As the role of talents rises, enterprises need to pay more attention to human capital. Meanwhile, intelligent production also urges intelligent HRM. Based on the accumulation of data, enterprises should use big data and artificial intelligence technologies to analyze employees, predict the future, and support businesses. Thus, the paper attempts to put forward ideas on data-driven solution to the promotion issue in HRM, and focus on the influence from the position of organization. The data come from a state-owned enterprise in China. Here the features of organizational position are chosen to analyze employee promotion and forecast employee prospect. From the analysis based on statistics and networks, as well as the prediction based on machine learning, we find that structural position plays a more critical role than geographic position. Besides, employees can benefit from working in the place where special experience is available, where mobility is more stable, or where resource is more abundant. But organizations should be concerned with the fair development of workers. The experimental results also validate that the prediction model is practical and effective.

[5] Title: Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming

Authors: DanaPessach

Description: In this paper, we propose a comprehensive analytics framework that can serve as a decision support tool for HR recruiters in real-world settings in order to improve hiring and placement decisions. The proposed framework follows two main phases: a local prediction scheme for recruitments' success at the level of a single job placement, and a mathematical model that provides a global recruitment optimization scheme for the organization, taking into account multilevel considerations. In the first phase, a key property of the proposed prediction approach is the interpretability of the machine learning (ML) model, which in this case is obtained by applying the Variable-Order Bayesian Network (VOBN) model to the recruitment data. Specifically, we used a uniquely large dataset that contains recruitment records of hundreds of thousands of employees over a decade and represents a wide range of heterogeneous populations. Our analysis shows that the VOBN model can provide both high accuracy and interpretability insights to HR professionals. Moreover, we show that using the interpretable VOBN can lead to unexpected and sometimes counter-intuitive insights that might otherwise be overlooked by recruiters who rely on conventional methods.

We demonstrate that it is feasible to predict the successful placement of a candidate in a specific position at a pre-hire stage and utilize predictions to devise a global optimization model. Our results show that in comparison to actual recruitment decisions, the devised framework is capable of providing a balanced recruitment plan while improving both diversity and recruitment success rates, despite the inherent trade-off between the two.

III. EXISTING SYSTEM:

The existing system also follows a traditional approach to Customer Relationship Management (CRM) System because it consists of a limited number of features in the dataset.

The employee's performance might sometimes not continue to contribute to the customer’s needs as an employee’s perspective focuses on the aspect that whether the “Project is built right?” whereas from a customer’s perspective it depicts whether the “Right Project is built?”. Hence, the employee only does the process of verification whereas the customer performs the validation task.

It has hence, been depicted that almost 50% of the employees do project verification but not validation and it has led to bad relations with customers hence losing many

Disadvantages:

- It is hard to predict the clusters formed
- Final results get impacted more with initial seeds
- Final results are impacted more by order of data

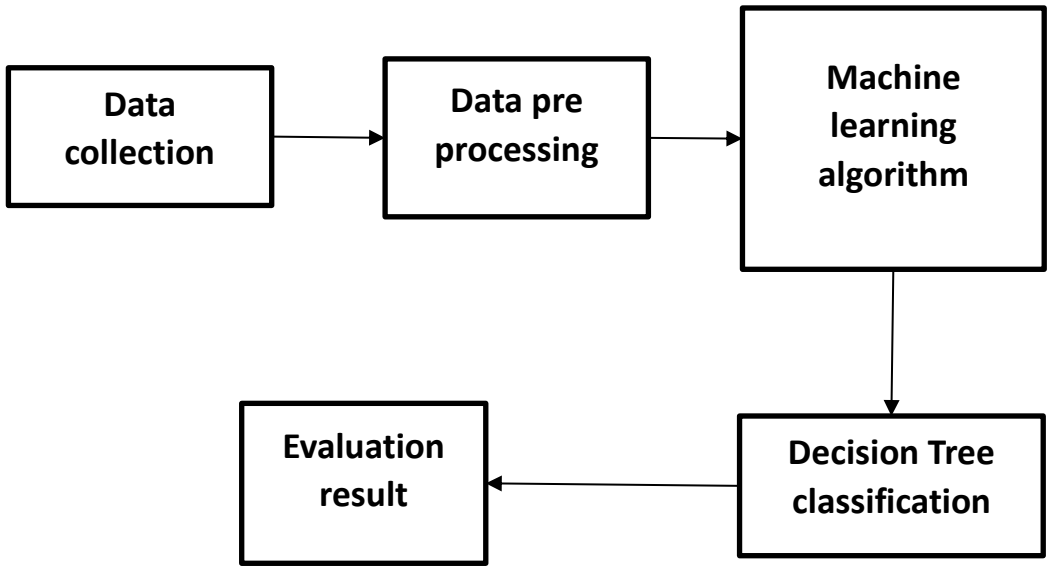
IV. PROPOSED SYSTEM

The system using additional features in the dataset has two benefits- it improves CRM within an organization and it improves employee performance analysis. This project uses a decision-tree based classification algorithm known as "machine learning" algorithm. Based on predicting the employee attrition rate and the employees emotional assessment in an organization. With the help of various machine learning algorithms this analysis is carried forward. This proposed system aims to predict the overall statistical employee performance analysis and feedback in addition to analyzing attributes such as Work from Home performance, etc. especially applicable to the current pandemic scenario.

Advantages

- Easy implementation
- Machine Learning is computationally faster with large number of variables and produces tighter clusters compared to hierarchical clustering
- If the centroids are re-calculated, this can change the entire cluster.

V. SYSTEM ARCHITECTURE



Module Description

- DATA COLLECTION
- DATA PRE-PROCESSING
- EDA concept
- MODEL IMPLEMENTATION
- PROMOTION OR NOT

MODULE 1: DATA COLLECTION

To collect the Dataset. Collecting data for training the ML model is the basic step in the machine learning pipeline. The predictions made by ML systems can only be as good as the data on which they have been trained. Following are some of the problems that can arise in data collection:

- Inaccurate data. The collected data could be unrelated to the problem statement.
- Missing data. Sub-data could be missing. That could take the form of empty values in columns or missing images for some class of prediction.
- Data imbalance. Some classes or categories in the data may have a disproportionately high or low number of corresponding samples. As a result, they risk being under-represented in the model.
- Data bias. Depending on how the data, subjects and labels themselves are chosen, the model could propagate inherent biases on gender, politics, age or region, for example. Data bias is difficult to detect and remove.

MODULE 2: DATA PRE-PROCESSING

Once the data is extracted from the twitter source as the datasets, this information has to be passed to the classifier. The classifier cleans the dataset by removing redundant data like stop words, emoticons in order to make sure that non textual content is identified and removed before the analysis.

Text pre-processing is an essential a part of any NLP method and the significance of the NLP pre-processing are

- To minimize indexing (or knowledge) records dimension of the textual content records
    1. Stop words bills 20-30% of total phrase counts in a special textual content record
    2. Stemming may just diminish indexing size as much as forty- 50%
- To make stronger the efficiency and effectiveness of the IR method
    1. Stop words aren't valuable for shopping or textual content mining
    2. Stemming used for matching the similar words in a text record

The sklearn.preprocessing package provides several common utility functions and transformer classes to change raw feature vectors into a representation that is more suitable for the downstream estimators.

In general, learning algorithms benefit from standardization of the data set. If some outliers are present in the set, robust scalers or transformers are more appropriate. The behaviors of the different scalers, transformers, and normalizers on a dataset containing marginal outliers is highlighted in Compare the effect of different scalers on data with outliers.

Standardization, or Mean removal and Variance Scaling

Standardization of datasets is a common requirement for many machine learning estimators implemented in scikit-learn; they might behave badly if the individual features do not more or less look like standard normally distributed data: Gaussian with zero mean and unit variance.

Scaling features to a range

In practice we often ignore the shape of the distribution and just transform the data to center it by removing the mean value of each feature, then scale it by dividing non-constant features by their standard deviation.

For instance, many elements used in the objective function of a learning algorithm (such as the RBF kernel of Support Vector Machines or the l1 and l2 regularizers of linear models) assume that all features are centered around zero and have variance in the same order. If a feature has a variance that is orders of magnitude larger than others, it might dominate the objective function and make the estimator unable to learn from other features correctly as expected.

An alternative standardization is scaling features to lie between a given minimum and maximum value, often between zero and one, or so that the maximum absolute value of each feature is scaled to unit size. This can be achieved using MinMaxScaler or MaxAbsScaler, respectively.

The motivation to use this scaling include robustness to very small standard deviations of features and preserving zero entries in sparse data.

MaxAbsScaler works in a very similar fashion, but scales in a way that the training data lies within the range [-1,1] by dividing through the largest maximum value in each feature. It is meant for data that is already centered at zero or sparse data.

Normalization

Normalization is the process of **scaling individual samples to have unit norm.** This process can be useful if you plan to use a quadratic form such as the dot-product or any other kernel to quantify the similarity of any pair of samples.

This assumption is the base of the Vector Space Model often used in text classification and clustering contexts.

MODULE 3: EDA concept (Exploratory data/Visualized Analysis)

Exploratory Data Analysis or (EDA) is understanding the data sets by summarizing their main characteristics often plotting them visually. This step is very important especially when we arrive at modelling the data in order to apply Machine learning. Plotting in EDA consists of Histograms, Box plot, Scatter plot and many more. It often takes much time to explore the data. Through the process of EDA, we can ask to define the problem statement or definition on our data set which is very important. So in this tutorial, we will explore the data and make it ready for modelling.

1. Importing the required libraries for EDA

The libraries that are used in order to perform EDA (Exploratory data analysis)

2. Loading the data into the data frame.

Loading the data into the pandas data frame is certainly one of the most important steps in EDA, as we can see that the value from the data set is comma-separated. One thing to remember in this step is that uploaded files will get deleted when this runtime is recycled.

3. Checking the types of data

The datatypes because sometimes the MSRP or the price of the car would be stored as a string or object, if in that case, we have to convert that string to the integer data only then we can plot the data via a graph.

4. Dropping irrelevant columns

This step is certainly needed in every EDA because sometimes there would be many columns that we never use in such cases dropping is the only solution. In this case, the columns such as Engine Fuel Type, Market Category, Vehicle style, Popularity, Number of doors, Vehicle Size doesn't make any sense to me so I just dropped for this instance**.**

5. Renaming the column

In this instance, most of the column names are very confusing to read, so I just tweaked their column names. This is a good approach it improves the readability of the data set.

6. Dropping the duplicate rows

This is often a handy thing to do because a huge data set as in this case contains more than 10, 000 rows often have some duplicate data which might be disturbing, so here I remove all the duplicate value from the data-set. For example prior to removing I had 11914 rows of data but after removing the duplicates 10925 data meaning that I had 989 of duplicate data.

7. Dropping the missing or null values.

An outlier is a point or set of points that are different from other points. This is mostly similar to the previous step but in here all the missing values are detected and are dropped later. Now, this is not a good approach to do so, because many people just replace the missing values with the mean or the average of that column.

8. Detecting Outlier

Sometimes they can be very high or very low. It's often a good idea to detect and remove the outliers. Because outliers are one of the primary reasons for resulting in a less accurate model. Hence it's a good idea to remove them. The outlier detection and

removing that I am going to perform is called IQR score technique. Often outliers can be seen with visualizations using a box plot.

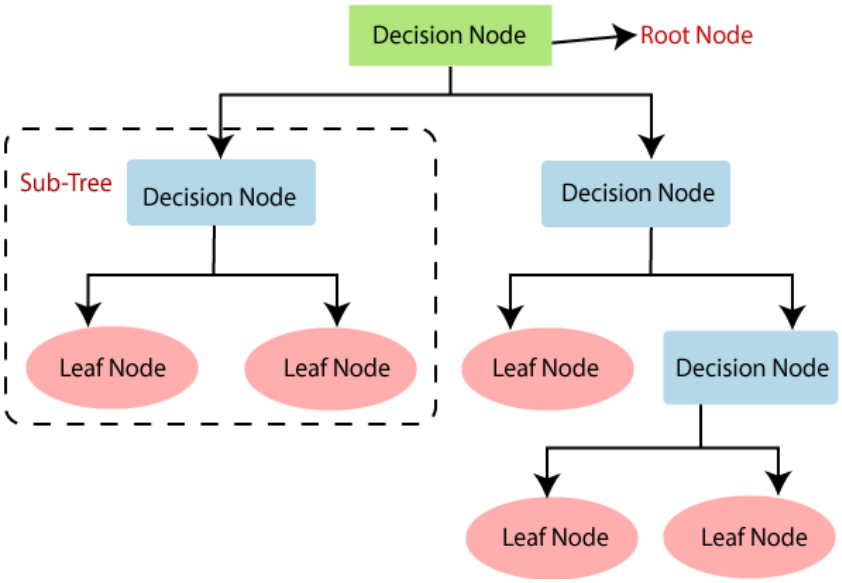
MODULE 4: MODEL IMPLEMENTATION

CLASSIFICATION ALGORITHM

- **DECISION TREE**

Decision trees are the building blocks of a random forest algorithm. A decision tree is a decision support technique that forms a tree-like structure. Each branch of the tree represents of possible decision occurrence or reaction.

A decision tree consists of three components: decision nodes, leaf nodes, and a root node. A decision tree algorithm divides a training dataset into branches, which further segregate into other branches. The nodes in the decision tree represent attributes that are used for predicting the outcome.



MODULE 5: PROMOTION OR NOT

Decision is regarded as a superior algorithm in terms of efficient memory utilization, high accuracy and low running times. It is simply highly robust and scalable technique to handle all sorts of noise from huge data sets and convert the data into a ready acceptable form for precision results. Hence the input is in csv format the output is displayed in 0's and 1's. where '0' is not promoted and '1' is promoted.

VI. CONCLUSION

During this research work, it has been identified that prediction of employee promotion is very important aspect in human capital building process. Three popular machine learning techniques for forecasting employee`s promotion were built and analyzed. Training dataset of employees were created using real data of companies and further processed. Popular supervised machine learning algorithms were studied and tested on performance using numerical metrics. The metrics, used in the research, showed that all three models were able to predict with close accuracy, despite small differences. To test the models for new samples, special software was developed, by using which it is possible to predict the advancement of employees. These tests showed the higher accuracy when using Decision Tree Classifier.

VII. FUTURE WORK

Since we have limited data to train the classifier, our approach is facing a high variance problem which can be observed in the learning curve as follows High variance problems can usually be mitigated by increasing the size of the dataset which should not be of much concern to Social Networks Organizations which already have fairly large datasets, In future researches, a new method will be presented; which can recognize the legitimate or fake account before any activity of the user in the network or at the time of registration. And we expect to run our model using more sophisticated concepts such as ontology engineering,

in order to semantically analyze user posts, and comportments. This later concept can improve the quality of prediction of fake or genuine profiles.

REFERENCES

[1] Yi, S., & Liu, X. (2020). Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. Complex & Intelligent Systems.

[2] Vohra, S., & Teraiya, J. (2013). A Comparative Study of Sentiment Analysis Techniques. International Journal of Information, Knowledge and Research in Computer Engineering, 2(2),313-317.

[3] Machine Learning & its Applications Outsource to India. (2020). Retrieved on May 18, 2020.

[4] Jain, A. P., & Dandannavar, P. (2016). Application of machine learning techniques to sentiment analysis. Second International Conference on Applied and Theoretical Computing and Communication Technology (ICATccT).

[5] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal.

[6] Aydogan, E., & Akcayol, M. A. (2016). A comprehensive survey for sentiment analysis tasks using machine learning techniques. 2016 International Symposium on Innovations in Intelligent Systems and Applications (INISTA).

[7] Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine learning techniques for sentiment analysis: A review. International journal of Multi-disciplinary science and Engineering.

[8] Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review.

[9] Kumar, A., & Sebastian, T. M. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. International Journal of Intelligent Systems and Applications.

[10] Swathi, R., & Seshadri, R. (2017). Systematic survey on evolution of machine learning for big data. International Conference on Intelligent Computing and Control Systems (ICICCS.

[11] Kawade, D., & Oza, K. (2017). Sentiment Analysis: Machine Learning Approach. International Journal of Engineering and Technology.

[12] Patil, H., & Atique, M. (2015). Sentiment Analysis for Social Media: A Survey. International Conference on Information Science and Security(ICISS).

[13] Yang, C., Tang, X., Wong, Y. C., & Wei, C.-P. (2010). Understanding Online Consumer Review Opinions with Sentiment Analysis using Machine Learning. Pacific Asia Journal of the Association for Information Systems.

[14] Gopu, M., & Swarnalatha, P. (2017). Analyzing customer sentiments using machine learning techniques. International Journal of Civil Engineering and Technology(IJCIET).

[15] Liu, R., SHI, Y., JI, C., & JIA, M. (2019). A Survey of Sentiment Analysis Based on TransferLearning. IEEE Access.

[16] Naiknaware, B., Kushwaha, B., & Kawathekar, S. (2017). Social Media Sentiment Analysis using Machine Learning Classifiers. International Journal of Computer Science and Mobile Computing.

[17] Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. Human-Centric Computing and Information Sciences.

[18] Yogi, T. N., & Paudel, N. (2020). Comparative Analysis of Machine Learning Based Classification Algorithms for Sentiment Analysis. International Journal of Innovative Science, Engineering & Technology.

[19] Mahendran, N., & Mekala, T. (2018). A survey: Sentiment analysis using machine learning techniques for social media analytics. International Journal of Pure and Applied Mathematics.