

# DATA SCIENCE PROCESS PIPELINE TO SOLVE EMPLOYEE ATTRITION AND THEIR JOB PERFORMANCE AND PREDICTING WITH AI

<sup>1</sup>MIR MUJAHED ALI KHAN, <sup>2</sup>MOHAMMED RAFATH ALI, <sup>3</sup>MOHAMMED AFZAL

<sup>1,2</sup>B.E STUDENT, <sup>3</sup>ASSOCIATE PROFESSOR

DEPARTMENT OF CSE

ISL Engineering College, Bandlaguda, Hyderabad

## Abstract:

Among all employee related problems, employee attrition is one of the key problem in the today's scenario despite the changes in the external environment. Attrition is said to be gradual reduction in number of employees through resignation, death and retirement. The other name given for Attrition is attrition. When a well-trained and well-adapted employee leaves the organization for any of the reason, it creates an empty space in an organization (i.e) there occurs a vacuum in the organization. It creates a great difficulty for a Human resource personel to fill the gap that has occurred. Modern Human resource managers is taking various steps to reduce the employee attrition rate and it has been a pivotal challenge for today's Managers. Many of the employees may also tend to leave the job for various undisclosed factors such as lack of job security, lack of career advancement, desire for change in new opportunities, anticipating higher pay, problems with supervisors and few other personal reasons. This study helps in predicting employee attrition.

## I. INTRODUCTION

Machine learning is a form of AI that enables a system to learn from data rather than through explicit programming. However, machine learning is not a simple process. As the algorithms ingest training data, it is then possible to produce more precise models based on that data. A machine learning model is the output generated when you train your

machine learning algorithm with data. After training, when you provide a model with an input, you will be given an output. For example, a predictive algorithm will create a predictive model. Then, when you provide the predictive model with data, you will receive a prediction based on the data that trained the model. Machine learning enables models to train on data sets before being deployed. Some machine- learning models are online and continuous. This iterative process of online models leads to an improvement in the types of associations made between data elements. Due to their complexity and size, these patterns and associations could have easily been overlooked by human observation. After a model has been trained, it can be used in real time to learn from data. The improvements in accuracy are a result of the training process and automation that are part of machine learning.

## II. SYSTEM ANALYSIS

### Existing System:

- Earlier the results show on basis of information-sharing behaviors, information-seeking behaviors and general usage behaviors during working hours
- Then shifted to some prediction analysis using Support vector machine, Naïve Bayes
- Drawback of this system was it was time consuming as well as

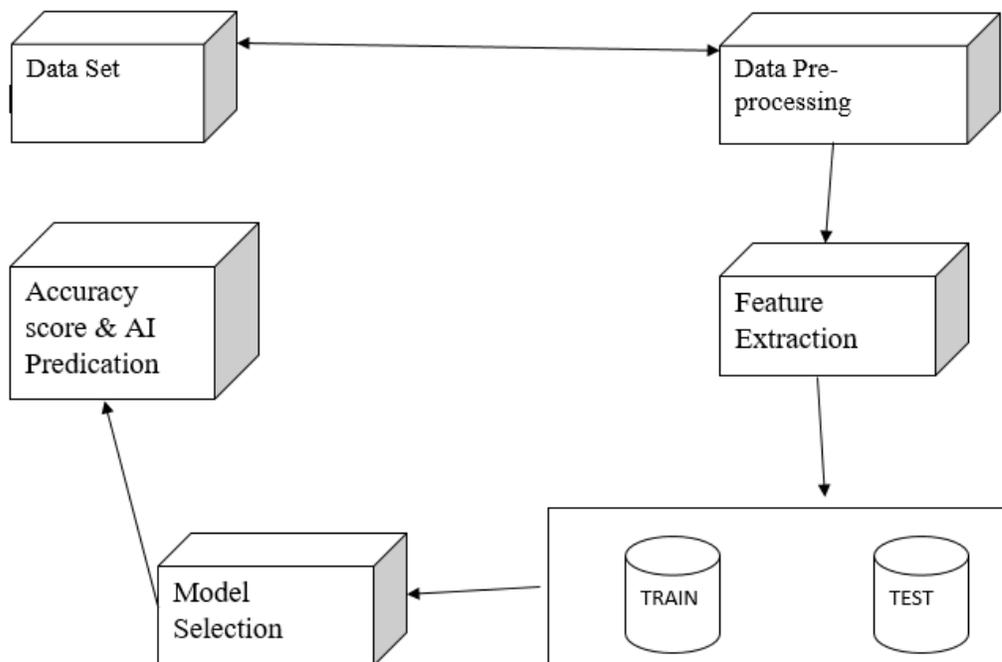
performance wise it is low, To increase the efficiency we moved to proposed methodology.

**Proposed System:**

- We know that larger companies contain more than thousand employees working for them, so taking care of the needs and satisfaction of each employee is a challenging task to do, it results in valuable and talented employees

leave the company without giving the proper reason. This project provides solution for the given problem as it gives a prediction model that can be used to predict which employee will leave the company and which will not leave. It also helps in finding the exact reasons which are motivating the employees for shifting companies like lower salary, less promotions or heavy work load etc. To find the result in the form of yes or no.

**System Architecture:**



**III. LITERATURE SURVEY:**

**Social media and the world of work: A strategic approach to employees' participation in social media**

**Author:** S. Dreher

The aim of this paper is to explore the risks and benefits of employees' social media use for an organization's reputation, and delivers suggestions for a strategic management approach. Design/methodology/approach –

The findings of this research paper are supported a comprehensive literature review and supported by a number one practice example. Findings – Through social media, employees function as powerful brand ambassadors who shape reputation with everything they are doing and say online. this needs a strategic management approach to employees' social media use, including research, internet access at the workplace, a robust commitment from the C-suite, the establishment of social media teams, the

implementation of guidelines and policies, training and education, integration, also as goal setting and measurement. These eight key steps will help communications professionals to raised prevent the risks and leverage the advantages of their employees' participation within the social web. Originality/value – This scientific research is made upon significant deficits within the management approach to employees' social media use and provides eight strategic key steps to raised handle employees' participation in social conversations.

**Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations.**

- **Author:** P. M. Leonardi, M. Huysman, and C. Steinfield

Social media are increasingly implemented in work organizations as tools for communication among employees. it's important that we develop an understanding of how they allow and constrain the communicative activities through which work is accomplished because it's these very dynamics that constitute and perpetuate organizations. we start by offering a definition of enterprise social media and providing a rough historical account of the varied avenues through which these technologies have entered and still enter the workplace. We also review areas of research covered by papers during this special issue and papers on enterprise social media published elsewhere to require stock of the present state of out knowledge and to propose directions for future research.

**Social networking, social network technologies, and the enterprise.**

**Author:** M. H. Jarrahi

This research proposal sets bent investigate the organizational consequences of Corporate Social Networking Technologies

(cSNT). Despite the rapid uptake of cSNT little or no is understood about the uses in and effects of SNT on enterprises. My main contention is that these technologies are embedded within a network of social ties. to know the interplay between the cSNT and therefore the informal network within organization, the proposed research will draw on different streams of literature, including the social network perspective, the knowledge sharing literature, and actor network theory. By building on these theoretical foundations, i will be able to discuss expected contributions of this research, and therefore the proposed research design.

**Motivations for social networking at work**

- **Author:** J. DiMicco, D. R. Millen, M. Geyer, C. Dugan, B. Brownholtz, and M. Müller

The introduction of a social networking site inside an outsized enterprise enables a replacement method of communication between colleagues, encouraging both personal and professional sharing inside the protected walls of a corporation intranet. Our analysis of user behavior and interviews presents the case that professionals use internal social networking to create stronger bonds with their weak ties and to succeed in bent employees they are doing not know. Their motivations in doing this include connecting on a private level with coworkers, advancing their career with the corporate , and campaigning for his or her projects. Author Keywords Social networking, enterprise, user motivation, CSCW

**IV. IMPLEMENTATION**

**DATA COLLECTION:**

- Data used in this paper is a set of employee details in the IBM records. This step is concerned with selecting the subset of all available data that

you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called *labelled data*.

#### **DATA PRE-PROCESSING:**

- Organize your selected data by formatting, cleaning and sampling from it.
- Three common data pre-processing steps are:
- Formatting
- Cleaning
- Sampling

#### **FORMATTING:**

The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.

#### **Cleaning:**

Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonymized or removed from the data entirely.

#### **Sampling:**

There may be far more selected data available than you need to work with. More data can result in much longer

running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

feature extraction

evaluation model

#### **ALGORITHM**

#### **RANDOM FOREST**

Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision *trees*, resulting in a *forest of trees*, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

#### **HOW RANDOM FOREST WORKS**

The following are the basic steps involved in performing the random forest algorithm

1. Pick N random records from the dataset.
2. Build a decision tree based on these N records.
3. Choose the number of trees you want in your algorithm and repeat steps 1 and 2.
4. For classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority vote.

**ADVANTAGES OF USING RANDOM FOREST**

pros of using random forest for classification and regression.

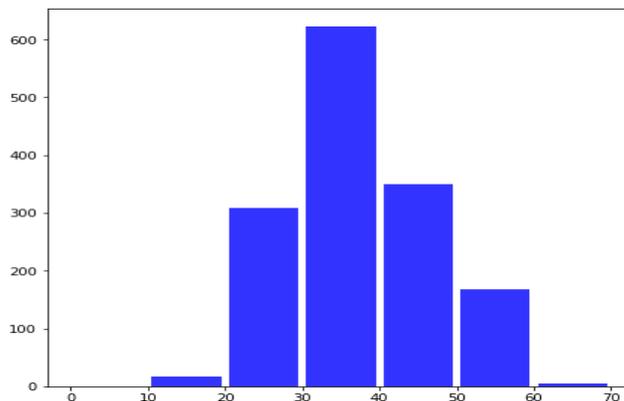
1. The random forest algorithm is not biased, since, there are multiple trees and each tree is trained on a subset of data. Basically, the random forest algorithm relies on the power of "the crowd"; therefore, the overall

biasedness of the algorithm is reduced.

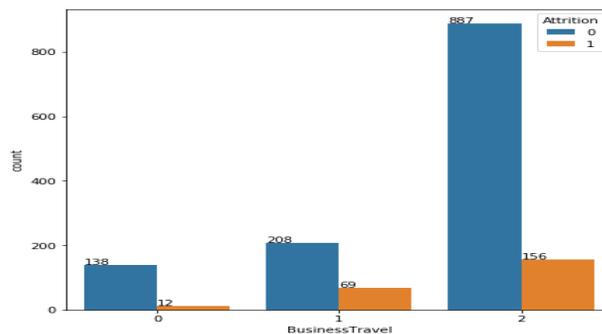
2. This algorithm is very stable. Even if a new data point is introduced in the dataset the overall algorithm is not affected much since new data may impact one tree, but it is very hard for it to impact all the trees.
3. The random forest algorithm works well when you have both categorical and numerical features.
4. The random forest algorithm also works well when data has missing values or it has not been scaled we.

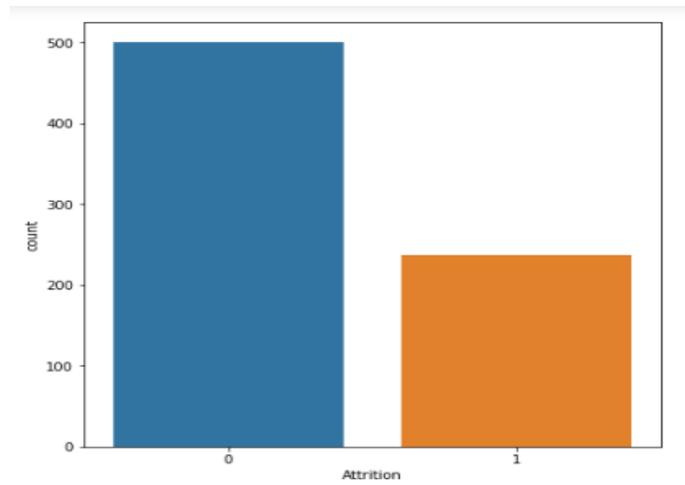
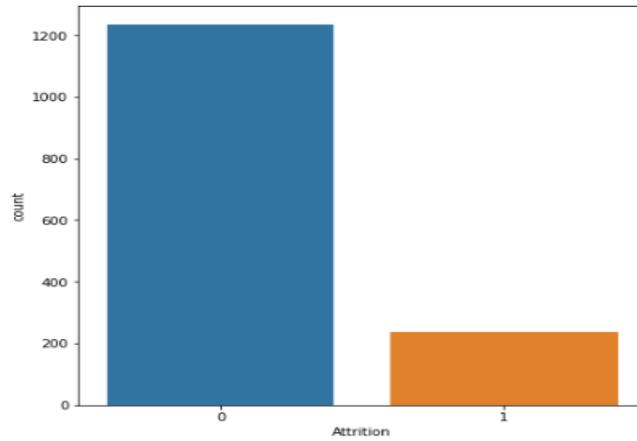
**V. SCREEN SHOT:**

**Data visualization with Histogram for a attribute**



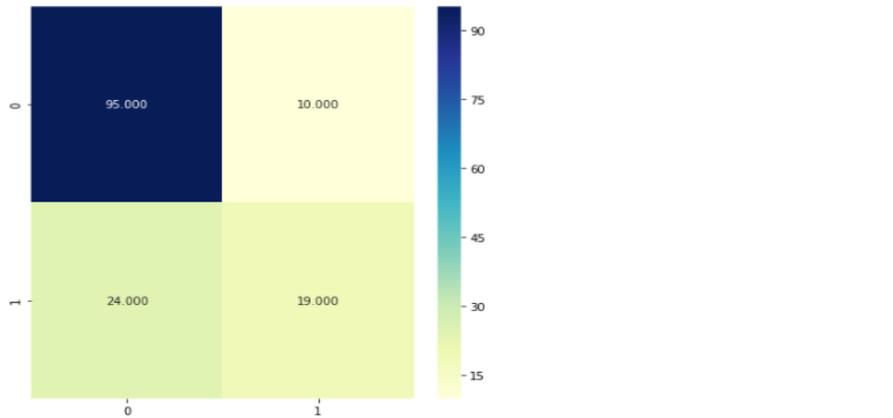
**Data visualization with Countplot**



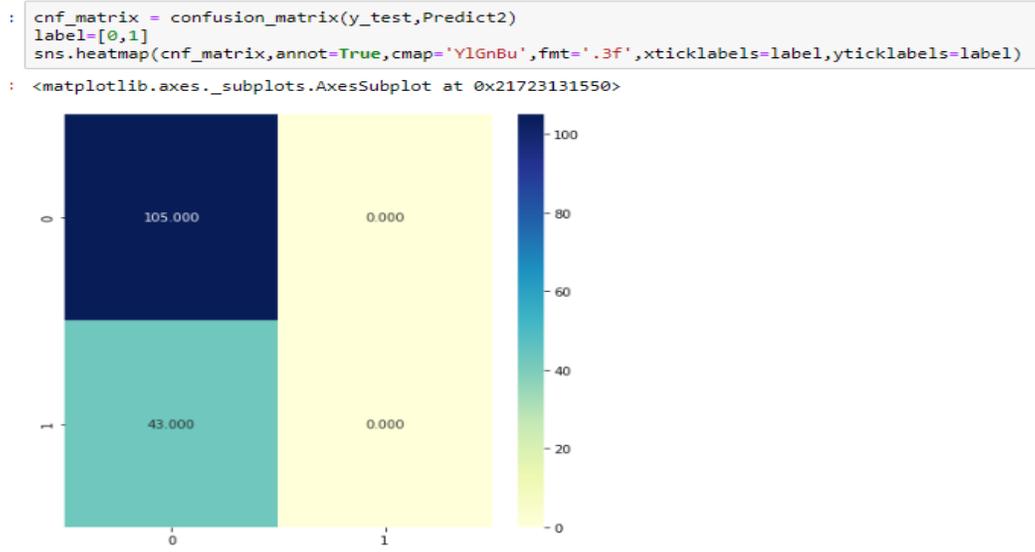


### Confusion Matrix with Random Forest

```
label=[0,1]  
sns.heatmap(cnf_matrix,annot=True,cmap='YlGnBu',fmt='.3f',xticklabels=label,yticklabels=label)  
<matplotlib.axes._subplots.AxesSubplot at 0x21723088438>
```



## Confusion Matrix with SVM



### VI. Conclusion:

The main objective of the organization is to earn profit and to earn profit the employer should concentrate in retaining talents and concentrate in making them stick to the organization for the long run. Employees are the assets of the organization. Hence it is important for the employers to minimize the attrition rate and help in both individual as well as organizational growth. Thus, Organizations should create an environment that fosters ample growth opportunities, appreciation for the work accomplished and a friendly cooperative atmosphere that makes an employee feel connected in every respect to the organization. Retention plans are an inexpensive way of enhancing workplace productivity and engaging employees emotionally. Proficient employees keep the quality up and business operations run smoothly along with the cost saving in the longer run.

### REFERENCES:

[1] S. Dreher, “Social media and the world of work: A strategic approach to employees’ participation in social media,” Corporate

Commun., Int. J., vol. 19, no. 4, pp. 344–356, Sep. 2014.

[2] A. P. McAfee, “Enterprise 2.0: The dawn of emergent collaboration,” MIT Sloan Manage. Rev., vol. 47, no. 3, pp. 21–28, 2006.

[3] P. M. Leonardi, M. Huysman, and C. Steinfield, “Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations,” J. Comput.-Mediated Commun., vol. 19, no. 1, pp. 1–19, Oct. 2013.

[4] M. H. Jarrahi, “Social networking, social network technologies, and the enterprise,” in Proc. iConference, Feb. 2011, pp. 688–690.

[5] M. Brandel, “Social networking behind the firewall,” Computer World, Tech. Rep., Aug. 2008. [Online]. Available: <https://www.computerworld.com/article/2551401/the-new-employee-connection-social-networking-behind-the-firewall.html>

[6] J. DiMicco, D. R. Millen, M. Geyer, C. Dugan, B. Brownholtz, and M. Müller, “Motivations for social networking at work,” in Proc. ACM Conf. Comput.

Supported Cooperat. Work, Nov. 2008, pp. 711–720.

[7] C. Steinfield, J. M. Dimicco, N. B. Ellison, and C. Lampe, “Bowling Online: Social networking and social capital within the organization,” in Proc. 4th Int. Conf. Commun. Technol., Jun. 2009, pp. 245–254.

[8] M. J. Brzozowski, “Water Cooler: Exploring an organization through enterprise social media,” in Proc. Int. Conf. Supporting Group Work, May 2009, pp. 219–228.

[9] Z. Mo and Y. Yu, “The mechanism of empowering leadership’s effect on employee creativity in dingtalk alibaba context,” in Proc. 4th Int. Forum Decision Sciences. Singapore: Springer, 2017, pp. 99–117.

[10] K. Le-Nguyen, Y. Guo, and J. Qiong, “Developing enterprise social software in China: Insights from Mingdao,” J. Inf. Technol. Teaching Cases, vol. 7, no. 2, pp. 62–78, 2017.

[11] J. Bughin, M. Chui, A. Miller, “How companies are benefiting from Web 2.0,” McKinsey Quart., vol. 9, no. 4, pp. 84–85, Jun. 2009.

[12] M. Quarterly. Business and Web 2.0: An Interactive Feature. Accessed: Apr. 2010. [Online]. Available: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/business-and-web20-an-interactive-feature>

[13] R. N. Landers and R. C. Callan, “Validation of the beneficial and harmful work-related social media behavioral taxonomies: Development of the work-related social media questionnaire,” Social Sci. Comput. Rev., vol. 32, no. 5, pp. 628–646, Mar. 2014.

[14] J. R. Carlson, R. B. Harris, K. J. Harris, and D. S. Carlson, “Social media use in the workplace: A study of dual effects,” J. Org.

End User Comput., vol. 28, no. 1, pp. 15–31, Jan. 2016.

[15] J. C. Y. Ng, I. Y. T. Shao, and Y. Liu, “This is not what I wanted: The effect of avoidance coping strategy on non-work-related social media use at the workplace,” Employee Relations, vol. 38, no. 4, pp. 466–486, Jun. 2016.

[16] R. M. Davison, C. X. J. Ou, and M. G. Martinsons, “Information technology to support informal knowledge sharing,” Inf. Syst. J., vol. 23, no. 1, pp. 89–109, Jan. 2013.