IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 10, October 2019

Human Resource Management: Big Data Analytics

Shruthi K Murthy[1], Chethan SR[2], Chethan K[3], Tejas SP[4]

Student, BE, Department of IEM, BIT, Bangalore, India[1, 2, 3]

Student, BE, Department of Civil Engineering, RVCE, Bangalore, India[4]

Abstract: Human Resource Management can be defined as planning, organizing, directing and compensating human resources resulting in the creation and development of human relation with a view to contribute proportionately to the organizational and individual goals. The examination of raw or crude data and drawing conclusions out of it is called data analytics. In this paper we will be analysing the employee turnover pattern and the factors contributing to it. Efforts will be made to create a model that can predict if a certain employee will leave the company or not. The goal is to create or improve different retention strategies on targeted employees. The first step in data analytics- data pre-processing is presented in the paper. Data pre-processing techniques convert crude data into useful format. Real world data are generally incomplete- noisy, inconsistent and contains many errors. Removing these factors improves the quality of analysis and prediction. The focus of data analytics lies in inference, the process of deriving conclusions. In this paper 2 out of top 3 strategies affecting employee turnover are being analysed and graphs plotted. The 3 top features include evaluation v/s exit, average monthly income v/s exit and satisfaction v/s exit.

Keywords: Human Resource Management, examination of raw or crude data and drawing conclusions- data analytics, employee turnover pattern, data pre-processing, evaluation v/s exit, average monthly income v/s exit, satisfaction v/s exit, planning, organizing, directing and compensating

I. INTRODUCTION

In this information era, huge amount of data is being stored, exchanged and conditioned. The volume of data that one has to deal with has exploded to unimaginable levels. Most of the data exists in its crude form and needs to be converted to useful format before analysis. This process of converting raw data into useful format is called data pre-processing. Real world data is [1]

- Incomplete: consists of missing attribute values or consists of only aggregate data.
- Noisy: containing errors or outliers.
- Inconsistent: containing discrepancies in code.

II. MOTIVATION

Encourage the commitment of employees to increase their performance and also be loyal to the organization as a whole. Emphasis on the quality of employees engaged in organizations goes a long way in producing quality goods and services, which is of great benefit both to the customers and the organization. Ensuring flexibility plays an important part in the way employees are organized, this makes them to be adaptive and receptive to all forms of changes in all aspects of their jobs such as work hours; working methods. Integrating organizational goals into strategic planning in order to make these policies cut across ranks and files of organization and ensuring that they are gladly accepted and implemented on daily routine by line managers. [2]

III. PROBLEM STATEMENT

In this paper we consider a company doing business in this 21st century world. This company has employed large number of employees working in different departments. The size of the company is huge and has several departments. The company is existent in business from a very long period of time and several employees have already left the company. The company's historical data is well tabulated and records maintained. The company wants to understand what factors contributed most to employee turnover and to create a model that can predict if a certain employee will leave the company or not. The goal is to create or improve different retention strategies on targeted employees.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 8, Issue 10, October 2019

```
In [4]:
#Importing Data
df = pd.read_csv(r'C:\Users\manasaav\datascience\employee-turnover-analysis\data.csv')
```

Figure 2 shows the Python code to import data from respective directory/ file.

```
In [3]:
import pandas as pd
import numpy as np
import seaborn as sns
%matplotlib inline
```

Figure 1 shows the Python code to import libraries.

IV. METHODOLOGY

A. Importing Libraries

Figure 1 shows the Python code to import libraries. We have used the following libraries: [3]

- NumPy is the fundamental package for scientific computing with Python.
- Pandas is for data manipulation and analysis. Panadas is an open source, BSD- licenced library providing easy-to-use data structures and data analysis tools.
- Matplotlib is a python 2D plotting library. It can be used in Python scripts, Jupyter notebook, web application servers and IPython shells.
- Seaborn is a Python data visualization library based on matplotlib for attractive and informative statistical graphics.

B. *Importing data:* Figure 2 shows the Python code to import data from respective directory/ file. The data stored in CSV format is being imported.

C. *Checking for missing values:* It is very essential in data pre-processing to check for missing values. Figure 3 shows the Python code to check for missing values. In this attempt no missing values were found.

D. *Renaming and rearranging the columns:* It is essential to rename the columns so that analysis is effective. Figure 4 shows the process of renaming the columns and figure 5 shows an effort to move the column 'exit' to the end as it has to be predicted.

E. *Exit rate:* Exit rate of the employees need to be checked. Figure 6 shows the exit ratio calculation. 76% of the employees stayed and 24% of employees exited.

```
In [5]:
#Checking whether our data contains any missing value or not
df.isnull().any()

Out[5]:
satisfaction_level       False
last_evaluation          False
number_project           False
average_montly_hours     False
time_spend_company       False
Work_accident            False
left                     False
promotion_last_5years    False
sales                    False
salary                   False
dtype: bool
```

Figure 3 shows the Python code to check for missing values.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 10, October 2019

V. EVALUATION V/S EXIT

- There is a bimodal distribution for those that had an exit. [4]
- Employees with low performance tend to leave the company more.
- Employees with high performance tend to leave the company more.
- The sweet spot for employees that stayed is within 0.6-0.8 evaluation. Figure 7 shows the employee evaluation distribution.

VI. AVERAGE MONTHLY HOURS V/S EXIT

- Another bimodal distribution for employees that exited.
- Employees who had less than 150 hours of work left the company more.
- Employees who had more than 250 hours of work left the company more. Figure 8 shows the plot of employee average monthly hours distribution.

In [7]:

```
#Renaming the columns
df = df.rename(columns={'satisfaction_level': 'Satisfaction',
                        'last_evaluation': 'Evaluation',
                        'number_project': 'ProjectCount',
                        'average_montly_hours': 'AverageMonthlyHours',
                        'time_spend_company': 'YearsAtCompany',
                        'Work_accident': 'WorkAccident',
                        'promotion_last_5years': 'Promotion',
                        'sales' : 'Department',
                        'left' : 'Exit'
                        })
df.head()
```

Out[7]:

Figure 4 shows the process of renaming the columns

In [6]:

```
#Moving the column 'Exit' to the end which is to be predicted
front = df['Exit']
df.insert(0, 'Exit', front)
df.head()
```

Out[6]:

Figure 5 shows an effort to move the column ‘exit’ to the end as it has to be predicted.

```
Exit_Rate = df.Exit.value_counts()/len(df)
Exit_Rate
```

Out[9]:

```
0    0.761917
1    0.238083
Name: Exit, dtype: float64
```

Figure 6 shows the exit ratio calculation.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 10, October 2019

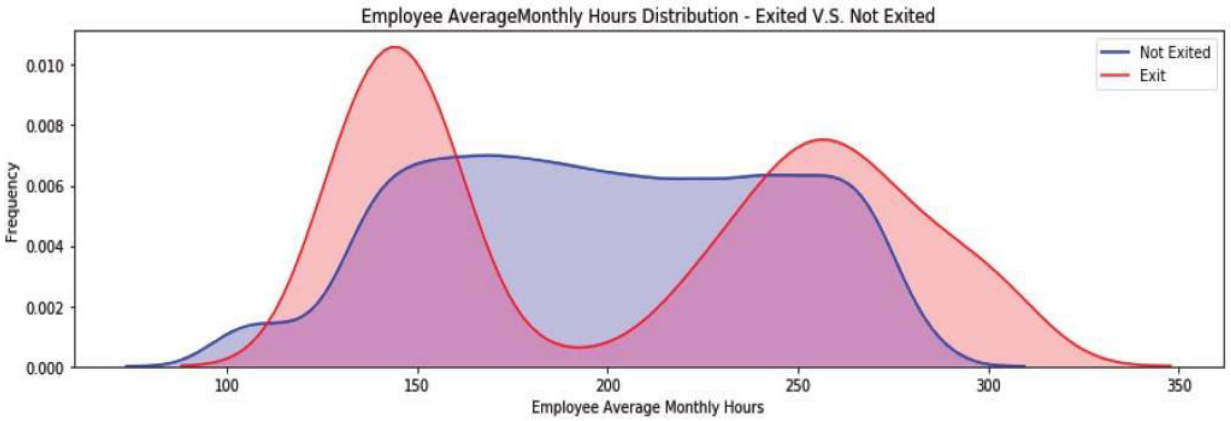

Figure 8 shows the plot of employee average monthly hours distribution.

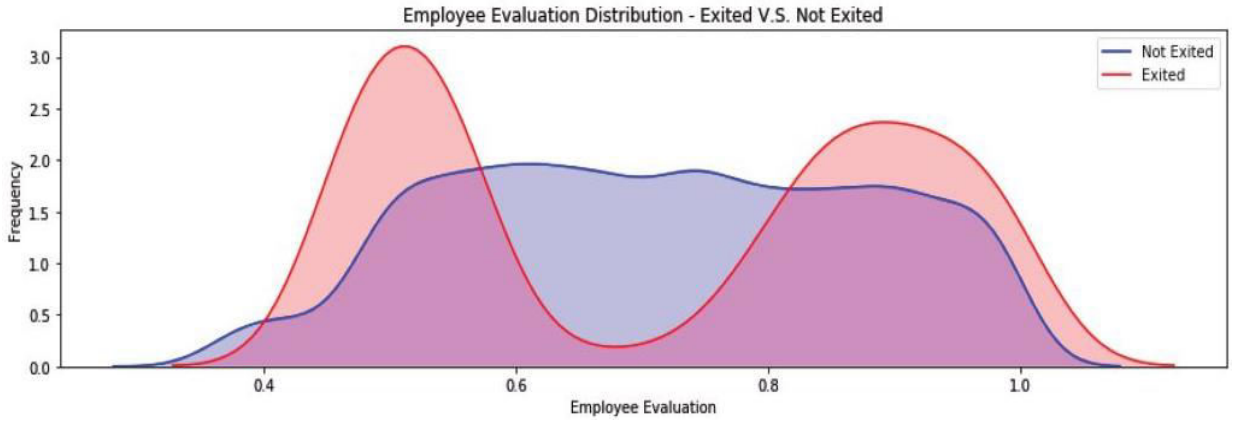

Figure 7 shows the employee evaluation distribution.

VII. CONCLUSION

The term "big data" can be applied to any of several different projects and technologies sharing the ultimate goal of supporting analysis on these large, heterogeneous, and evolving data sets. The term "data science" refers to the statistical, technical, and domain-specific knowledge required to ensure that the analysis is done properly. Techniques for managing some common causes for bad data and invalid analysis have been used in other areas, such as data warehousing and distributed database. However, big data projects face special challenges when trying to combine big data and data science without producing inaccurate, misleading, or invalid results. [5][6] This paper discusses potential causes for "bad big data science", focusing primarily on the data quality of the input data, and suggests methods for minimizing them based on techniques originally developed for data warehousing and distributed database projects. A company proactive in business in this 21st century world had many workers leaving the company. Data analytics had to be carried out on the data –both historical and present trend to draw inference. The goal was to create or improve different retention strategies on targeted employees working in different departments of the company. A python code was written and executed in the Jupyter platform to analyse and draw conclusions. The first step in data analytics- data pre-processing was successfully carried out and exit ratio calculated. 2 out of top 3 strategies affecting employee turnover are being analysed and graphs plotted. The 3 top features include evaluation v/s exit, average monthly income v/s exit and satisfaction v/s exit.

REFERENCES

[1]. Principles of data mining- DJ Hand - Drug safety, 2007 – Springer
[2]. Human Resource Management: Theory and Practice- July 2012- ISBN- 978-978-50666-8-5
[3]. Python for Data Analysis- Python data science libraries- Boston University- https://pdfs.semanticscholar.org/407c/05a87c5eb47af7e7cddf414a23d2b4dfdac1.pdf
[4]. Experimental data and statistical models for bimodal EM failures- 2000 IEEE International Reliability Physics Symposium Proceedings. 38th Annual (Cat. No.00CH37059)- DOI: 10.1109/RELPHY.2000.843940
[5]. The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness Published in: 2015 IEEE International Conference on Big Data (Big Data)- DOI: 10.1109/BigData.2015.7363988
[6]. Bad Big Data Science- 2016 IEEE International Conference on Big Data (Big Data)- DOI: 10.1109/BigData.2016.7840935

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 10, October 2019

BIOGRAHY

VISHESH S (BE, MBA, PGDIB, (MTech)) born on 13th June 1992 hails from Bangalore (Karnataka) and has completed B.E in Telecommunication Engineering from VTU, Belgaum, Karnataka in 2015. He also worked as an intern under Dr. Shivananju BN, former Research Scholar, Department of Instrumentation, IISc, Bangalore. His research interests include Embedded Systems, Wireless Communication, BAN and Medical Electronics. He is also the Founder and Managing Director of the corporate company Konigtronics Private Limited. He has guided over a thousand students/interns/professionals in their research work and projects. He is also the co-author of many International Research Papers. He has recently completed his MBA in e-Business and PG Diploma in International Business. Presently Konigtronics Private Limited has extended its services in the field of Software Engineering and Webpage Designing. Konigtronics also conducts technical and non-technical workshops on various topics. Real estate activities are also carried out under the guidance of Siddesh B S BE (civil). Vishesh S along with his father BS Siddesh has received various awards and applauses from the scientific and entrepreneurial society. He was appointed as the MD of Konigtronics Pvt Ltd (INC. on 9th Jan 2017) at an age of 23 years. His name is indexed in various leading newspapers, magazines, scientific journals and leading websites & entrepreneurial forums. He is also the guide for many international students pursuing their Masters.