Predicting Employee Attrition using Random Forest with IBM Cloud

1. Introduction

1.1 Overview :-

Employee attrition is a gradual but deliberate reduction in the number of employees in a company or business organization. Employees will at some point in time look to change their job places for a number of reasons. It might be for professional or personal reasons but it does happen.

But when the attrition rate starts to grow past some certain figures or threshold, then you have reasons to be concerned. Attrition may lead to a leadership gap in your organization if your senior leaders are the ones leaving or it can affect the diversity at your workplace if the minority groups are the ones leaving. Employee attrition can differ among organizations based on the kinds of people leaving but the definition of attrition remains the same.

To define attrition in simple terms, it is when there is a diminishing in the number of the workforce due to several factors that are very avoidable. Some of the reasons for employee attrition are a lack of confidence in the leadership and market value of the company, a work environment that is not conducive, absence of professional growth, etc.

1.2 Purpose :-

Every year a lot of companies hire a number of employees. The companies invest time and money in training those employees, not just this but there are training programs within the companies for their existing employees as well.

The concept of these programs is to increase the effectiveness of their employees. If there is any model that can predict employee attrition it will be easy to analyze the attrition process and improving the performance of employees.

The intension is to build a model that predicts the Attrition of the Employees based on the given factors of an employee using Machine Learning.

2. Literature Survey :-

2.1 Existing problem :-

Thus, we can see that till date there has been various theoretical and technical research and studies have been carried out to find the attrition prediction. But there has been no significant study or research on development of tool which can take automated decisions on categorizing valuable employees and ordinary employees. And there is no application that shows the final dashboard that shows the retention factors which HR Managers must consider while retaining the valuable employee; so that the Human Resource Management budget can be reduced significantly if retention rate is increased.

2.2. Proposed solution:-

Every year a lot of companies hire a number of employees. The companies invest time and money in training those employees, not just this but there are training programs within the companies for their existing employees as well. The concept of these programs is to increase the effectiveness of their employees. If there is any model that can predict employee attrition it will be easy to analyze the attrition process and improving the performance of employees.

The intensions is to build a model that predicts the Attrition of the Employees based on the given factors of an employee using Machine Learning.

3. THEORITICAL ANALYSIS:-

3.1. Block diagram:-

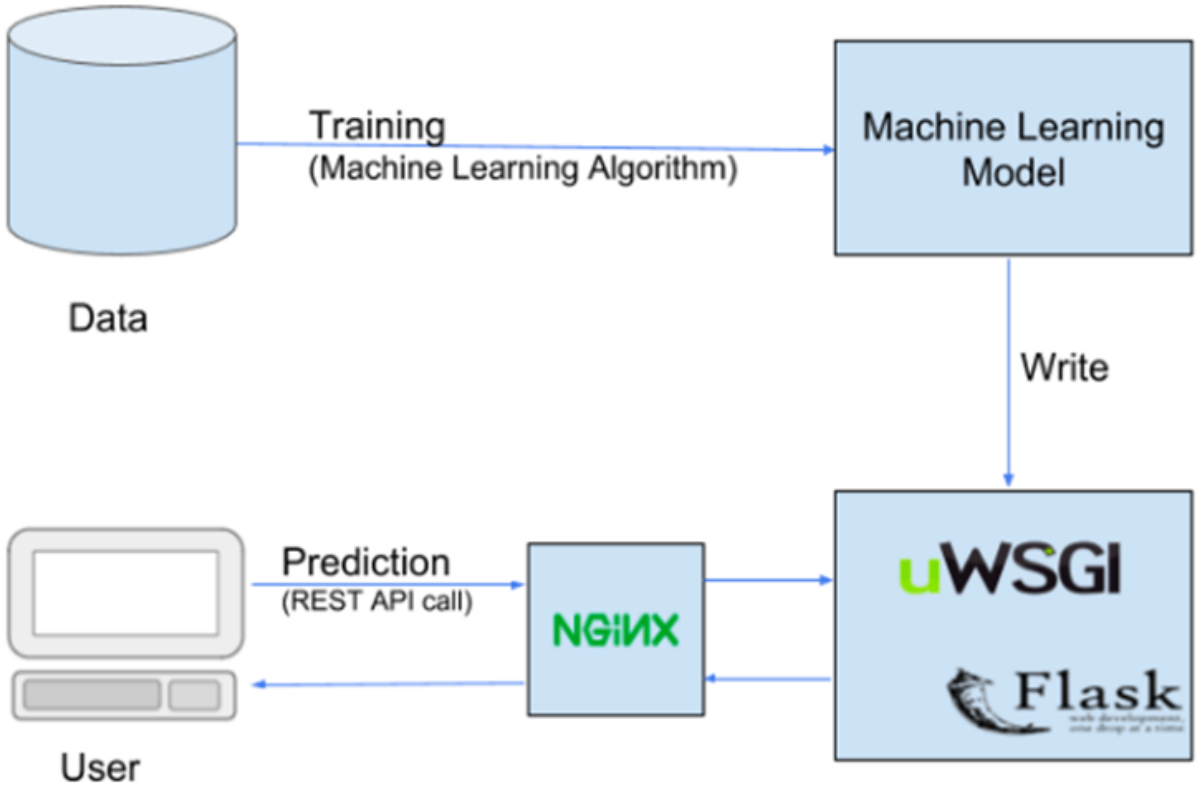


Fig:3.1.1 Block Diagram

3.2. Hardware/software designing :-

Software Requirements:-

- OS – Windows XP,7,8,10
- Jupyter Software
- Spyder Software
- Anaconda Command Prompt

Hardware Components:-

- Processor – i3
- Hard Disk Storage – 10 GB
- RAM – 1GB

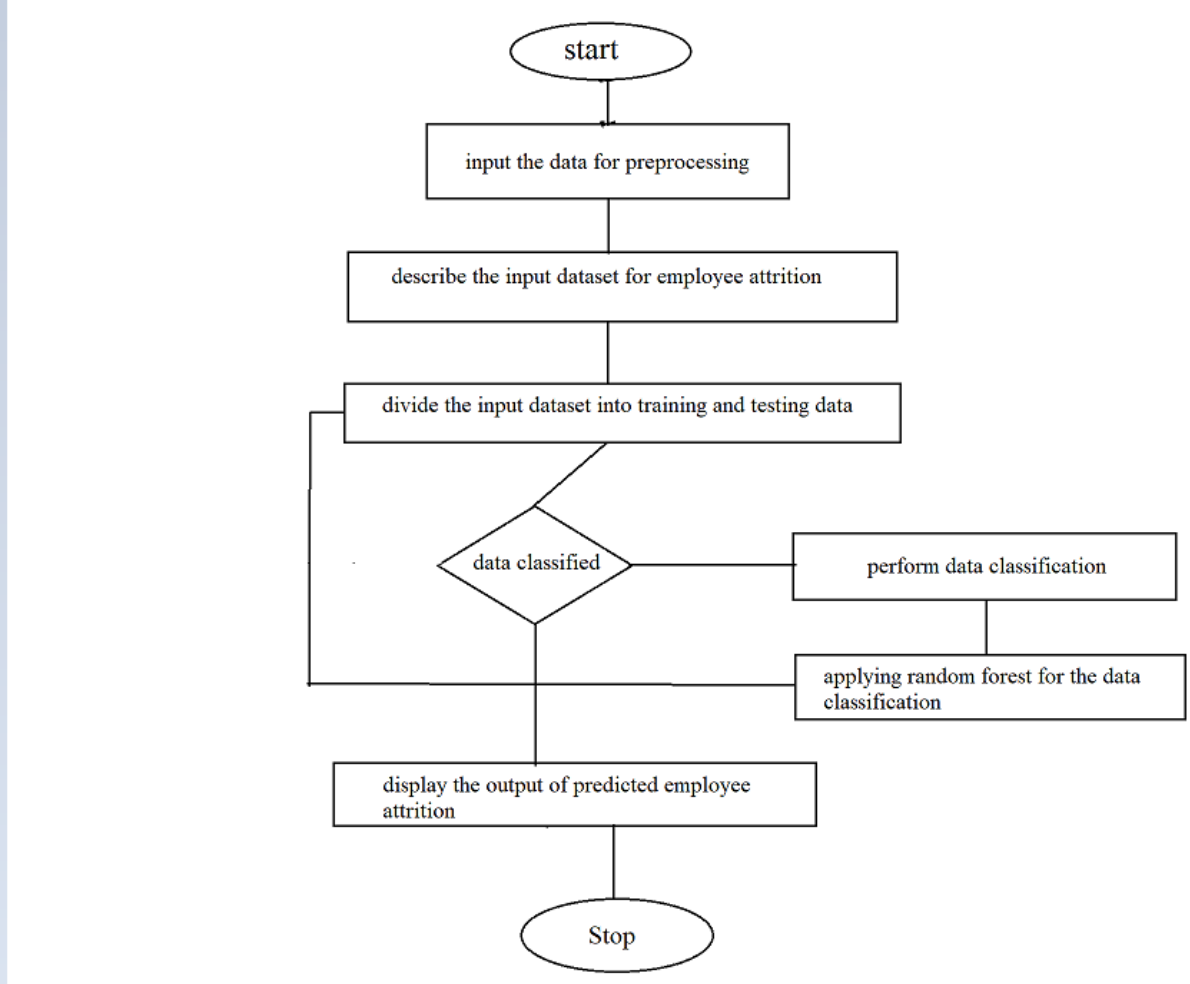
4. EXPERIMENTAL INVESTIGATIONS:-

In this section, the two simulated human resources datasets are used to assess the performance of our predictive models. The first one is the large sized-dataset supplied by Kaggle that contains 15000 samples where its target variable is ”left” and its 9 features are satisfaction level; last evaluation; number project; average monthly hours; time spend company; Work accident; promotion last 5 years; sales and Salary. The second simulated human resources analytics dataset is a medium sized-dataset provided by IBM and it contains 1470 samples with 34 features and its target variable is attrition that is represented as ”low” (employee did not left) or ”high” (employee left). In this second simulated dataset, we find our 11 selected features as part of its 34 features, so we will check the performance of our predictors using the entire dataset of IBM with its 34 features. Then, we will assess their performance using the same dataset but we will keep only the 11 selected features of our employee attrition model (Marital status, Age, Tenure, Grade, Rewards, Job involvement, Training, Business Travel, Job satisfaction, Job performance, and Environment satisfaction).

Table shows the results in terms of accuracy (that is defined as the percentage of the correctly classified data by the model and it represents the ratio of the predictions total number that is correct) and F1-score using the two simulated datasets.

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	sales	salary
0	0.38	0.53	2	157	3	0	1	0	sales	low
1	0.80	0.86	5	262	6	0	1	0	sales	medium
2	0.11	0.88	7	272	4	0	1	0	sales	medium
3	0.72	0.87	5	223	5	0	1	0	sales	low
4	0.37	0.52	2	159	3	0	1	0	sales	low

5.Flow Chart:-



6.Results:-

- This Employee attrition Prediction process constitutes of three steps. These steps are mentioned below:
- Input are taken from the user related to the Employee attrition prediction which includes satisfaction , last evalution, number of projects worked ,average monthly hours,time spend number of years worked at company,work accidents,promotions in last 5 years ,departments ,salary level used.
- In the next step we use the prediction by train and testing the data which is analyzed during the process of accuracy calculation.
- The output is generated related to the analysis based on the data collected related to the prediction of Employee attrition on various combinations.

7.ADVANTAGES AND DISADVANTAGES:-

Advantages:-

Not all turnovers are negative, we generally feel that an employee leaving the organization is detrimental to the organization, but there is a flip side to it.

Employees leaving an organization may lead to benefits. This type of job attrition is called '**healthy attrition**' and is needed for growth and development of an organization

1. Higher manpower costs:

There are times when employees stay with the organization for long, which might mean that they are getting top of their pay scale.

This means that these employees are being paid a lot more than others who are doing similar job but are comparatively new in the process.

This excessive manpower costs leads to financial burden which is generally not identified on a regular run of the companies.

2. Negative effect of people:

There are some employees in an organization who work against the culture of the organization and even affect the working environment.Such employees even go to an extend to impact the loyalty of others and their outlook towards the organization.

This could mean they are working against the organization from inside. When such employees leave the organization they lead to more profit than loss.

3. New idea:

Many a time when some people leave an organization they open gates for new talent and new ideas.

Mostly employees who are in the organization get used to the working atmosphere and get complacent. This means lack of risk and definitely no new ideas. When they leave there is space for people who have high risk taking caliber and stop the firm from becoming stagnant.

4. Higher performance:

There are employees who just have been working on a slow pace for years within an organization. They are reasons for poor performance and slow growth.

When they leave the organization the team becomes fast paced and the turnover time is decreased considerably. This means reduced cost and satisfied clients.

5. Setting the culture right:

There are times that employees do not care for a lot of organization's policies which creates a brand for the organization.

When strong actions are taken against such employees like asking them to leave because they were affecting the brand name in a negative way would lead to a strong message and create a professional culture in an organization.

Disadvantages:-

When employees leave the organization it is a loss to the company, the team and the individuals.

Employees are the backbone of any organization and their departing may lead to lot of various losses to company on different aspects. The disadvantages can be:

1. Decreased overall performance:

The whole business process is affected when an employee leave the organization. It is even more risky when this happen all of a sudden.

2. Daily task management:

Sudden attrition may lead to difficulty in managing daily tasks. Even large organization struggle to manage their task when employees leave jobs, getting small information and managing daily tasks become difficult as they cannot be managed by small current team which is left behind.

3. Increased cost:

This has to be the highest disadvantage to a company when employees leave their jobs. There is increased cost associated with every level of the process – **losing**

and paying the previous employee, hiring a new one, training cost for the new employee.

4. Lack of knowledgeable employees:

This goes without saying when employees leave an organization they take with them the experience they have gained overtime.

5. Create a Negative image:

It is not just that employees are looking for job, even organizations are on the outlook of qualified professionals. When any company has high attrition rate it negatively impact the brand of the organization.

8.APPLICATIONS:-

- Software compaines
- Government office
- Schools and Colleges
- Banking
- Hospitals
- Factories

9. CONCLUSION

The main goal of this research is to help HR managers to detect as soon as possible an employee's intention to leave using predictive analytics methods and so to fight this attrition. The contributions can be summarized into three points:

i) The proposal of a new employee attrition model that contains only 11 features necessary and sufficient to detect intention to leave and to predict positive attrition using a mixed research methodology.

ii) The proposal of machine, deep and ensemble learning predictive models and their experimentation in a variety of different settings (large-sized simulated dataset, medium sized simulated dataset and small-sized real dataset) to best assess their performance.

iii)The interpretation and the explication that enables HR managers to understand what makes an employee want to leave and to help them in adopting key policies to retention.

10.FUTURE SCOPE

In terms of study limitations, considering dynamic features that deal with employees' behaviour and their emotional states will be promising to study their impact on employee attrition. In this case, the predictive models training must be on-line as data will be dynamic and new data can be added whenever required. We acknowledge also that our questionnaire respondents have equally suggested other features to be considered and that can cause voluntary turnover and so can be integrated into our future study. In fact, they have proposed to consider health issues, job security and the use of new technologies in the company. Finally, in future research, considering unbalanced data is a real challenge especially for organizations and companies with high turnover rate because the adopted predictive models are experimentally not suitable for unbalanced data.

11. BIBLIOGRAPHY

1. Alex Frye, C. B. M. S., 2018. Employee Attrition: What Makes an Employee Quit?. SMU Data Science Review, 1(1).

2. Alf Crossman, B. A., 2003. Job satisfaction and employee performance of Lebanese banking staff. Journal of Managerial Psychology, 18(4), pp. 368-376.

3. Ananya Sarker, S. S. D. M. S. Z. M. M. R., 2018. Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm. Global Journal of Computer Science and Technology, 18(1), pp. 1-5.

.4. Beynon, M. J. P. P. D. a. P. G., 2015. nvestigating the impact of training influence on employee retention in small and medium enterprises: a regression-type classification and ranking believe simplex analysis on sparse data.. Expert Systems , Issue 32(1), pp. 141- 154.

5. Edouard Ribes, K. T. B. P., 2017. Employee turnover prediction and retention policies design: a case study, US: Cornell university. 23. Fabian Pedregosa, G. V. A. G., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 8/11(10/11).

6. Firth, L. M. D. M. K. A. a. L. C., 2004. How can managers reduce employee intention to quit?. Journal of managerial psychology, 19(2), pp. 170-187.

7. Gupta, S. S., 2010. Employee Attrition and Retention: Exploring the Dimensions in the urban centric BPO Industry, A-10, SECTOR 62, NOIDA, INDIA: JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA.

8. Hausknecht, J. P. R. J. M. &. H. M. J., 2008. Targeted Employee Retention: PerformanceBased and Job-Related Differences in Reported Reasons for Staying.. Working Paper Series, Cornell University, School of Industrial and Labour Relations, Centre for Advanced Human Resource Studies, Working Paper, Issue 08 – 06..

9. IBM , 2018. Employee Attrition and Performance - IBM Analytics. [Online] Available at: https://www.ibm.com/communities/analytics/watson-analytics-blog/hremployee-attrition/ [Accessed June 2018].

33. IBM Inc., 2017. HR Analytics. [Online] Available at: www-01.ibm.com/software/analytics/solutions/operational-analytics/hranalytics/ [Accessed 2017 Dec].

34. IBM Watson, 2018. Kenexa IBM Talent management. [Online] Available at: https://www.ibm.com/talent-management/case-studies/ibm-talentmanagement-case-studies [Accessed June 2018].

12. APPENDIX

a.SOURCE CODE

IndexEA.html:

```
<html lang="en" dir="ltr">

  <head>
    <title>Employee Attrition Predictor</title>
  </head>
<style>
.container{
width:100%;
text-align:center;
}
.ml-container{
width:100%;
text-align:center;
}

input{
font-size:1.0em;
width:20%;
text-align:center;
}
input placeholder{
text-align:center;
}
button {
  background-color: #f44336; /* Red */
  border: none;
  color: white;
  padding: 15px 32px;
  text-align: center;
  text-decoration: none;
  display: inline-block;
  font-size: 16px;
  margin: 4px 2px;
```

```
  cursor: pointer;
}
</style>
 <body>
<style>
body
{
background-image:url('../static/img.jpg');
background-position: center;
font-family:serif;
background-size:cover;
}
</style>


      <!-- Website Title -->
          <div class="container">
         <h2 class='container-heading' style="color:white;"><span class="heading_font">Employee
Attrition Predictor</span></h2>
         <div class='description'>
                        <p style="color:white;">A Machine Learning Web App, Built with Flask</p>
                </div>
          </div>

      <!-- Text Area -->
          <div class="ml-container">
                <form action="{{ url_for('predict') }}" method="POST">


<input class="form-input" type="text" name="satisfaction_level" placeholder="Enter satisfaction
level(0-1)"><br><br>
<input class="form-input" type="text" name="last_evaluation" placeholder="Enter last
evaluation(0-1)"><br><br>
<input class="form-input" type="text" name="number_project" placeholder="Enter number of
projects worked"><br><br>
<input class="form-input" type="text" name="average_montly_hours" placeholder="Enter average
montly hours"><br><br>

<input class="form-input" type="text" name="time_spend_company" placeholder="Enter number
of years worked at company"><br><br>
```

```
<select id="Work_accident" name="Any work accidents?">
<option value="">Any work accidents?</option>
   <option value="1">Yes</option>
   <option value="0">No</option>
 </select><br><br>

<select id="promotion_last_5years" name="Promotion in last 5years">
<option value="">Promotion in last 5years</option>
   <option value="1">Yes</option>
   <option value="0">No</option>
 </select><br><br>
<select id="department" name="Select your department">
   <option value="">Select your department</option>
   <option value="7">technical</option>
<option value="6">sales</option>
<option value="5">product_mng</option>
<option value="4">marketing</option>
<option value="0">RandD</option>
<option value="2">HR</option>
<option value="1">accounting</option>
<option value="3">management</option>
 </select><br><br>

<select id="salary" name="Select salary level">
   <option value="">Select your salary Level</option>
   <option value="1">Low</option>
<option value="2">Medium</option>
   <option value="0">High</option>

 </select><br><br>




                <button type="submit" class="my-cta-button">Predict</button>
        </form>

   </body>
</html>
```

ResultEA.html:

```
<html lang="en" dir="ltr">
	<head>
		<title style="color:white;">Employee Attrition Predictor</title>
	</head>
<style>
.container{
width:100%;
text-align:center;
}
.results{
width:100%;
text-align:center;
}
</style>
  <body>
<style>
body
{
background-image:url('../static/img.jpg');
background-position: center;
font-family:serif;
background-size:cover;
}
</style>
    <!-- Website Title -->
		<div class="container">
      <h2 class='container-heading' style="color:white;" ><span class="heading_font">Employee
Attrition Prediction</span></h2>
	<div class='description'>
			<p style="color:white;>A Machine Learning Web App using Flask.</p>
		</div>
	 </div>

		<!-- Result -->
		<div class="results">
			{% if prediction_text==1 %}
```

```
                                <h1 style="color:white;">Prediction: <span class='danger'>Chance
of employee leaving is high. Take immediate measures to save the spot.</span></h1>
                                <img class="gif" src="{{ url_for('static', filename='high.webp') }}"
alt="High Image">
                        {% elif prediction_text==0 %}
                                <h1 style="color:white;">Prediction: <span
class='safe'>Hooray!!!Chance of employee leaving is low!</span></h1>
                                <img class="gif1" src="{{ url_for('static', filename='low.webp') }}"
alt="Low Image">
                        {% endif %}
                </div>



        </body>
</html>
```

app.py:

```
# -*- coding: utf-8 -*-
"""
Created on Tue Sep 15 14:11:40 2020

@author: rincy
"""



from flask import Flask,request,render_template
import os
import numpy as np
import pandas as pd
import pickle
app = Flask(__name__)
model = pickle.load(open('PAE_model.pkl', 'rb'))
@app.route('/')
def home():
    return render_template("indexEA.html")
@app.route('/predict',methods=["POST","GET"])
```

```
def predict():
    input_features = [float(x) for x in request.form.values()]
    features_value = [np.array(input_features)]
    print(features_value)

    features_name = ['satisfaction_level', 'last_evaluation', 'number_project',
       'average_montly_hours', 'time_spend_company', 'Work_accident',
       'promotion_last_5years', 'department', 'salary']
    payload_scoring = {"input_data": [{"field": [['satisfaction_level', 'last_evaluation',
'number_project','average_montly_hours', 'time_spend_company', 'Work_accident','left',
'promotion_last_5years', 'department', 'salary']],
                       "values":  features_name}]}
    response_scoring =
requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/da786aa6-424e-
4a07-bdd2-6b51e95d497e/predictions?version=2021-07-12', json=payload_scoring,
headers={'Authorization': 'Bearer ' + mltoken})
    print("Scoring response")
    predictions = response_scoring.json()
    print(predictions)
    scaler=pickle.load(open("scaler.pkl","rb"))
    X_test_scaled=scaler.transform(features_value)
    prediction = model.predict(X_test_scaled)
    output=prediction[0]

    return render_template('resultEA.html', prediction_text=output)
if __name__=="__main__":
     port = int(os.getenv('PORT', 8000))
     app.run(host='0.0.0.0', port=port, debug=True)
      #http_server = WSGIServer(('0.0.0.0', port), app)
     #http_server.serve_forever()
```

b.UI screenshot output