PROJECT REPORT DOCUMENTATION

DATE	19-NOV-2022
TEAM ID	PNT2022TMID07232
PROJECT NAME	Corporate Employee Attrition Analytics

1.INTRODUCTION:

1.1 PROJECT OVERVIEW:

This project involves overall approach taken to create the attrition risk model, process flows, data streams involved and the output attained from the model. It further aims to recommend the thrust areas and best practices on employee retention at different stages of the employee's association with an organization.

1.2 PURPOSE:

Employee turnover has turn out to be a large venture for data technological know-how companies. The departure of key software program builders would possibly reason large loss an IT business enterprise in view that they additionally leave with essential commercial enterprise understanding and integral technical skills. It is fundamental for IT companies to apprehend developer turnover in order to keep certified builders and reduce injury due to developer exit. In this research, monthly self-report of the software developers includes developer's activities, working hours, no of projects they have been assigned etc. will been taken into account for analysis for doing the prediction with the help of data science algorithm. By the usage of NB algorithm, KNN algorithm and SVM algorithm, prediction mannequin has been in contrast on the experimental groundwork and supply the end result of which algorithm is performing better. Then, this fantastic mannequin will be given to HR managers to predict whether or not the worker will depart the corporation or not.

2.LITERATURE SURVEY:

2.1 EXISTING PROBLEM:

Employees are the treasured property of any organization. However, if you quit your job unexpectedly, the company will cost a lot of money. Not only are new employees wasting money and time, but new employees are also spending time making profits for their companies. Employee turnover is a diversity of existing staff and is replaced by new staff for a period of time. Staff turnover (Lingfeng Bao et all.,2017) is one of the most urgent troubles in the enterprise's body of workers management. Previous lookup on the difficulty of workforce turnover failed to be successful in managing, supervising, and stopping employee turnover on manufacturing traces with excessive turnover rates. Employee churn (Andry Alamsyah et all., 2018) has a range of bad penalties for a business, such as unequal workload distribution, huge monetary losses, and the extra time required to recruit a replacement, all of which can lead to

an expand in consumer unhappiness. Tangible fees encompass coaching costs and the time it takes from when a worker begins to when they begin contributing intangible expenses contain what is closing when an environment friendly worker quits: new product ideas, high-quality challenge management, or client relationships. A survey of 1,000 full-time employees performed by way of the on-line recruitment company (Tara Safavi et all., 2018) Headhunter.net reviews that 78% would take a new function if the proper possibility comes alongside and 48% these that are employed are on the lookout for sparkling opportunities. Because software program engineers may go away with a lot of crucial data and expertise, if developer turnover (Mingfei Teng et all.,2019) is no longer properly managed, it can undermine the success of a software program venture and end result in big losses for the firm. As a result, being in a position to pick out who will go away the organization early will permit the corporation to preserve brilliant software program builders whilst minimizing the loss when they leave.

2.2 REFERENCES:

1. Lingfeng Bao, Zhenchang Xing, “Who Will Leave the Company? A Large-Scale Industry Study of Develop.

2. Andry Alamsyah; Nisrina Salma, “A Comparative Study of Employee Churn Prediction Model", IEEE International Conference on Science and Technology, 2018.

3. Dilip Singh Sisodia, " Evaluation of Machine Learning Models for Employee Churn Prediction ", IEEE International Conference on Inventive Computing and Informatics, 2017.

4. Francesca Fallucchi, “Predicting Employee Attrition Using Machine Learning Techniques", Mdpi Journal of computers, 2020.

5. Mohammad Nayeem Hasan, “A Comparison of Logistic Regression and Linear Discriminant Analysis in Predicting of Female Students Attrition from School in Bangladesh”, International Conference on Electrical Information and Communication Technology,2019.

2.3 PROBLEM STATEMENT DEFINITION:

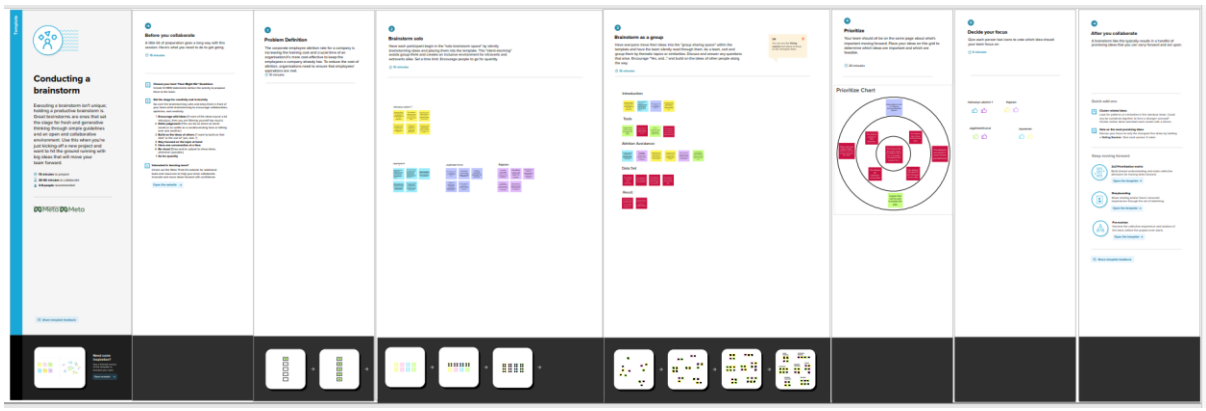
In today’s world many industries and specially IT are experiencing high attrition rate. Some common causes of attrition in their organization are known to managers and HR departments. This difficulty is triggered by using disappointment with various elements of a job, such as profession aspirations, work location, salary, overall performance management, job satisfaction, and managers, amongst others. Employee attrition (Francesca Fallucchi et all.,2020) manipulate is integral to the long-term fitness and success of any organization. Employees who leave on their own accord have a negative influence (James M. Vardaman et

all., 2015) on the organization or project in which they are employed. Any industry's HR and senior management, as well as policymakers, are collaborating to reduce voluntary exits.

3. IDEATION & PROPOSED SOLUTION:

3.1 EMPATHY MAP:

3.2 IDEATION & BRAINSTROMING:



3.3 PROBLEM SOLUTION:

Proposed Solution Template: Project team shall fill the following information in proposed solution template.

S No	Parameter	Description
1	Problem Statement (Problem to be solved)	Corporate Employee Attrition Analysis - How to retain employees effectively
2	Idea / Solution description	Prioritize the professional growth & give the pleasant workspace and use some classification algorithm to predict their retention and manage their relationship using this software.
3	Novelty / Uniqueness	Employee attrition prediction is specifically focused on identifying why employees voluntarily leave, what might have prevented them from leaving, and how we can use data to predict attrition risk.
4	Social Impact / Customer Satisfaction	Employee's attrition has huge impact on company, recruiting new employees and investing time to train them is increased. Losing a good employee creates a negative impact of profit on the company.
5	Business Model (Revenue Model)	The business is struggling with employee attrition. This software will be helpful to analyze the workforce trends and find the root cause of Attrition.
6	Scalability of the Solution	The dashboard is scalable for the companies when their employee's dataset is used for analysis. The model can successfully predict the futuristic approach and suggests preventive measures.

3.4 PROBLEM SOLUTION FIT:

Project Title: Corporate Employee Attrition Analytics

Project Design Phase-I - Solution Fit Template

Team ID: PNT2022TMID07232

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

Who is your customer?
i.e. working parents of 0-5 y.o. kids

- CEO of an organization.
- Shareholder of the organization.
- Human Resource person in a organization

6. CUSTOMER CONSTRAINTS CC

What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.

- Budget problem.
- Loss of experienced employees.
- Addressing the problem of all employees.

5. AVAILABLE SOLUTIONS AS

Which solutions are available to the customers when they face the problem
or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking

- Heard about the issues of employee's but unable to solve since it's groundless.
- Took survey's but unable to analyze the surveys.

Solution:
- Better analysis is provided to find the problems.

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

2. JOBS-TO-BE-DONE / PROBLEMS J&P

Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.

- Providing reasonable leaves.
- Recognize and reward employees.
- Create learning and development programs.
- Prioritize employee's happiness.
- Offer flexibility

9. PROBLEM ROOT CAUSE RC

What is the real reason that this problem exists?
What is the back story behind the need to do this job?
i.e. customers have to do it because of the change in regulations.

- Poor work environment.
- Work pressure.
- Poor rewards and no recognisation.
- Employee have to do it because of change in rules and regulations that are not employee friendly.

7. BEHAVIOUR BE

What does your customer do to address the problem and get the job done?
i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)

- Did surveys, and found some problems of employees.
- Spend time with employees to find major problem.
- Providing support sessions.

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

3. TRIGGERS TR

What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.

Maintaining company's standard.
When the company's profit decrease.
Less productive activities of employees.

4. EMOTIONS: BEFORE / AFTER EM

How do customers feel when they face a problem or a job and afterwards?
i.e. lost, insecure > confident, in control - use it in your communication strategy & design.

- Lost.
- Depressed.
- Loss of confidence.

10. YOUR SOLUTION SL

If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.
If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.

Solution:
- Doing proper analysis and identifying major problems.
- Providing clear and simple solution to solve every problem.

8. CHANNELS of BEHAVIOUR CH

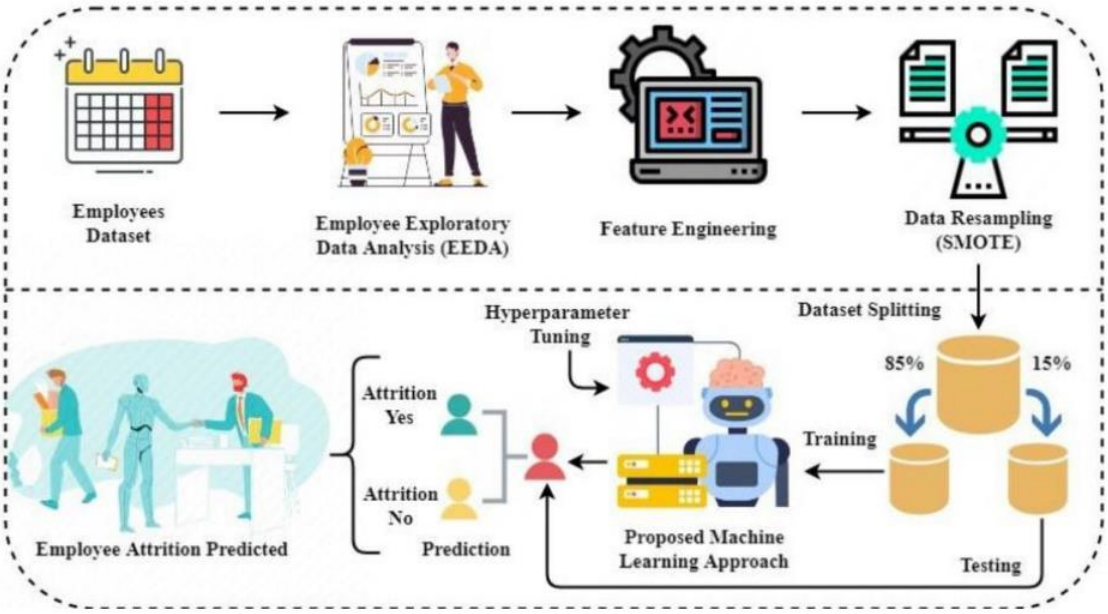
8.1 ONLINE
What kind of actions do customers take online? Extract online channels from #7

8.2 OFFLINE
What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.

Online:
Did surveys, and found some problems of employees.
Offline:
Spend time with employees to find major problem.
Providing support sessions.

Identify strong TR & EM

Solution Architecture



4. FUNCTIONAL REQUIREMENTS:

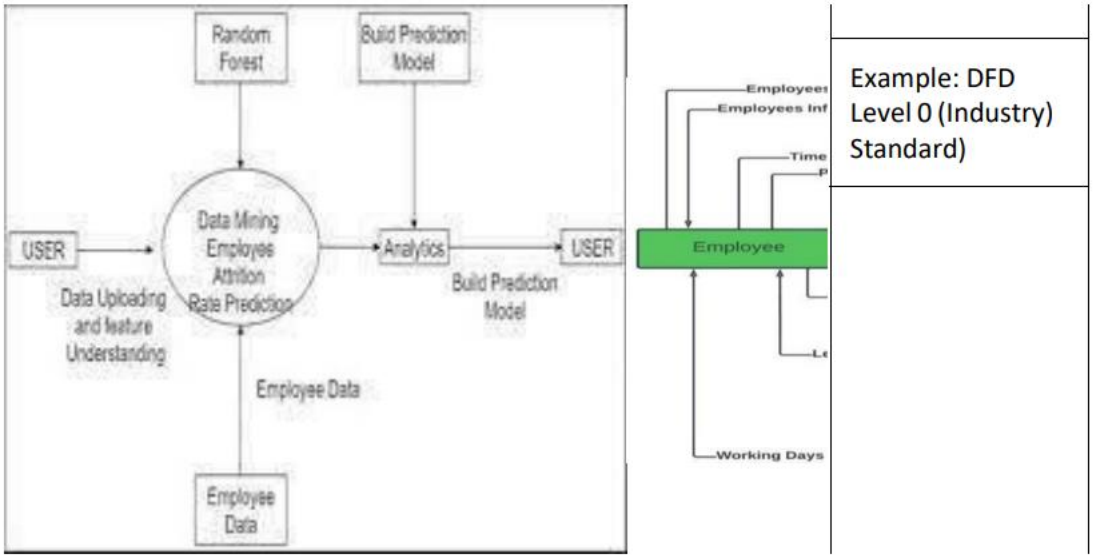
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIn
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP

FR-3	Account Creation	Create an account in the Profile Dashboard
FR-4	Input Credentials	Uploading your dataset Analyzing the attrition rate using dashboard
FR-5	Processing Methods	Using IBM Cognos Analytics Dashboard Using Prediction algorithm to find attrition rate

5. PROJECT DESIGN:

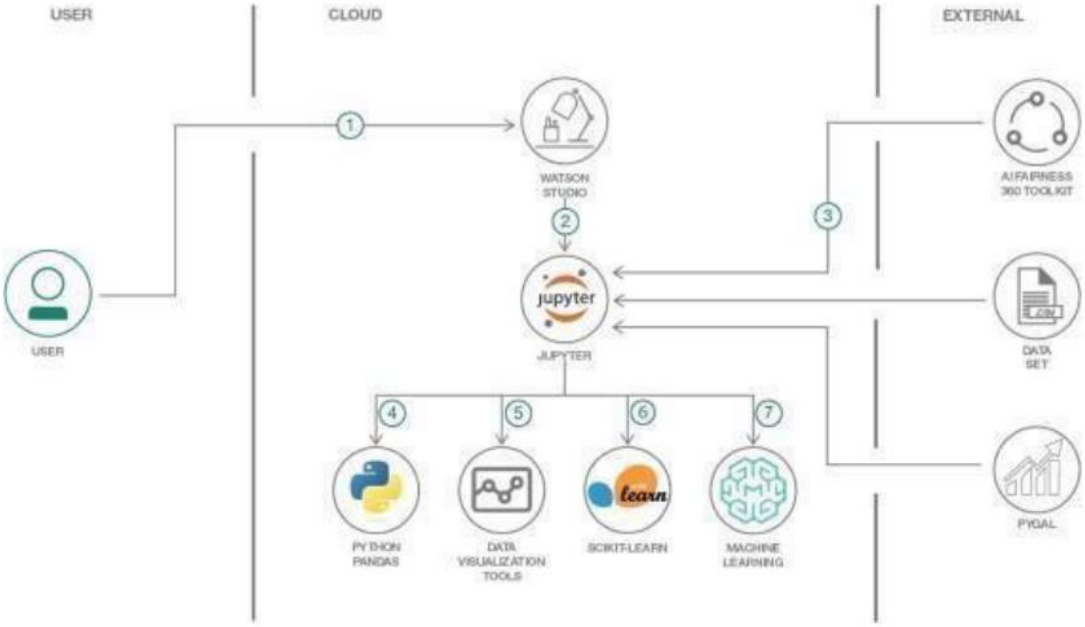
5.1 DATA FLOW DIAGRAMS:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 TECHNICAL ARCHITECTURE:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2.



5.3 USER STORIES:

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Corporate workers)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through my mail	I can register & access the dashboard with Gmail Login	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard					
Administrator	Login	USN-1	As an admin, I can login using email id and password			
	Data collection	USN-2	As an admin, I can load the data for visualization			
	Visualization	USN-3	The data can be visualized in various forms and used for analyzing.			

6. PROJECT PLANNING & SCHEDULING:

6.1 PLANNING AND ESTIMATION:

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring technical papers,research publications etc.	28 SEPTEMBER 2022
Prepare Empathy Map	Prepare empathy map canva to capture the user pains & gains, prepare list of problem statements	24 SEPTEMBER 2022
Ideation	List the idea by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	25 SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document which includes the novelty, feasibility of idea, business model, social impact, scalability of solution etc.	23 SEPTEMBER 2022

6.2 SPRINT DELIVERY SCHEDULE:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Dashboard	USN-1	I provide the information about the employees who work for our company as a user for the attrition detail.	5	High	AishwaryaLakshmi, Jeyaraman K, Rajaram, Jagathesh Kumar
Sprint-1		USN-2	I will examine the dataset as an Analyst and carry out exploratory data analysis in Cognos Analytics.	3	High	AishwaryaLakshmi, Jeyaraman K, Rajaram, Jagathesh Kumar
Sprint-2	Report	USN-3	I want minimal, simpler visualisations that report certain events as a user.	2	Low	AishwaryaLakshmi, Jeyaraman K, Rajaram, Jagathesh Kumar
Sprint-2		USN-4	I'll create a report using Cognos Analytics as an analyst.	3	Medium	AishwaryaLakshmi, Jeyaraman K, Rajaram, Jagathesh Kumar

7. CODING & SOLUTIONING:

7.1 FEATURE 1:

Importing the necessary files, libraries for plotting and the algorithms from sklearn.

In [ ]:
```
import math, time, random, datetime

# data analysis and wrangling
import pandas as pd
import numpy as np
from pandas_profiling import ProfileReport
```

In [ ]:
```
# visualization
import seaborn as sns
import matplotlib.pyplot as plt
plt.style.use('seaborn-whitegrid')

#import for interactive plotting
import plotly.offline as py
py.init_notebook_mode(connected=True)
import plotly.graph_objs as go
import plotly.tools as tls
import plotly.figure_factory as ff
from plotly.subplots import make_subplots
%matplotlib inline
```

In [ ]:
```
# Preprocessing
from sklearn.preprocessing import OneHotEncoder, LabelEncoder, label_binarize, StandardScaler
```

In [ ]:
```
# Preprocessing
from sklearn.preprocessing import OneHotEncoder, LabelEncoder, label_binarize, StandardScaler
```

In [ ]:
```
# machine learning
from sklearn import model_selection, tree, preprocessing, metrics, linear_model
from sklearn.metrics import confusion_matrix,classification_report
from sklearn.svm import SVC, LinearSVC
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.linear_model import Perceptron,SGDClassifier,LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split,StratifiedKFold, GridSearchCV, learning_curve, cross_val_score
#from catboost import CatBoostClassifier, Pool, cv
```

In [ ]:
```
# ignore Warnings
import warnings
warnings.filterwarnings('ignore')
```

In [ ]:
```
df = pd.read_csv("/content/Attrition_prediction.csv")
```

In [ ]:
```
df.head()
```

Out[ ]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	Stan
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	

5 rows × 35 columns

7.2 FEATURE 2:

The performance accuracy of various algorithms used for analysing.

In [ ]:
```
# Function that runs the requested algorithm and returns the accuracy metrics
def fit_ml_algo(algo, X_train,y_train, cv):

    # One Pass
    model = algo.fit(X_train, y_train)
    acc = round(model.score(X_train, y_train) * 100, 2)

    # Cross Validation
    train_pred = model_selection.cross_val_predict(algo,X_train,y_train,cv=cv,n_jobs = -1)

    # Cross-validation accuracy metric
    acc_cv = round(metrics.accuracy_score(y_train, train_pred) * 100, 2)

    return train_pred, acc, acc_cv
```

In [ ]:
```
# Logistic Regression
start_time = time.time()
train_pred_log, acc_log, acc_cv_log = fit_ml_algo(LogisticRegression(), X_train,y_train, 10)
log_time = (time.time() - start_time)
print("Accuracy: %s" % acc_log)
print("Accuracy CV 10-Fold: %s" % acc_cv_log)
print("Running Time: %s" % datetime.timedelta(seconds=log_time))
```

```
Accuracy: 89.8
Accuracy CV 10-Fold: 88.63
Running Time: 0:00:01.753627
```

In [ ]:
```
# SVC
start_time = time.time()
train_pred_svc, acc_svc, acc_cv_svc = fit_ml_algo(SVC(),X_train,y_train,10)
svc_time = (time.time() - start_time)
print("Accuracy: %s" % acc_svc)
print("Accuracy CV 10-Fold: %s" % acc_cv_svc)
print("Running Time: %s" % datetime.timedelta(seconds=svc_time))
```

```
Accuracy: 88.53
Accuracy CV 10-Fold: 85.91
Running Time: 0:00:00.497278
```

In [ ]:
```
# Linear SVC
start_time = time.time()
train_pred_svc, acc_linear_svc, acc_cv_linear_svc = fit_ml_algo(LinearSVC(),X_train, y_train,10)
linear_svc_time = (time.time() - start_time)
print("Accuracy: %s" % acc_linear_svc)
print("Accuracy CV 10-Fold: %s" % acc_cv_linear_svc)
print("Running Time: %s" % datetime.timedelta(seconds=linear_svc_time))
```

```
Accuracy: 89.89
Accuracy CV 10-Fold: 88.73
Running Time: 0:00:01.055932
```

In [ ]:
```
# K Nearest Neighbour
start_time = time.time()
train_pred_knn, acc_knn, acc_cv_knn = fit_ml_algo(KNeighborsClassifier(n_neighbors = 3),X_train,y_train,10)
knn_time = (time.time() - start_time)
print("Accuracy: %s" % acc_knn)
print("Accuracy CV 10-Fold: %s" % acc_cv_knn)
print("Running Time: %s" % datetime.timedelta(seconds=knn_time))
```

```
Accuracy: 88.82
```

In [ ]:
```
# Gaussian Naive Bayes
start_time = time.time()
train_pred_gaussian, acc_gaussian, acc_cv_gaussian = fit_ml_algo(GaussianNB(),X_train,y_train,10)
gaussian_time = (time.time() - start_time)
print("Accuracy: %s" % acc_gaussian)
print("Accuracy CV 10-Fold: %s" % acc_cv_gaussian)
print("Running Time: %s" % datetime.timedelta(seconds=gaussian_time))
```

```
Accuracy: 68.9
Accuracy CV 10-Fold: 66.28
Running Time: 0:00:00.198038
```

In [ ]:
```
# Perceptron
start_time = time.time()
train_pred_gaussian, acc_perceptron, acc_cv_perceptron = fit_ml_algo(Perceptron(),X_train,y_train,10)
perceptron_time = (time.time() - start_time)
print("Accuracy: %s" % acc_perceptron)
print("Accuracy CV 10-Fold: %s" % acc_cv_perceptron)
print("Running Time: %s" % datetime.timedelta(seconds=perceptron_time))
```

```
Accuracy: 81.24
Accuracy CV 10-Fold: 83.58
Running Time: 0:00:00.310640
```

In [ ]:
```
# Stochastic Gradient Descent
start_time = time.time()
train_pred_sgd, acc_sgd, acc_cv_sgd = fit_ml_algo(SGDClassifier(),X_train, y_train,10)
sgd_time = (time.time() - start_time)
print("Accuracy: %s" % acc_sgd)
print("Accuracy CV 10-Fold: %s" % acc_cv_sgd)
print("Running Time: %s" % datetime.timedelta(seconds=sgd_time))
```

```
Accuracy: 86.59
Accuracy CV 10-Fold: 84.74
```

In [ ]:
```
# Decision Tree
start_time = time.time()
train_pred_dt, acc_dt, acc_cv_dt = fit_ml_algo(DecisionTreeClassifier(),X_train, y_train,10)
dt_time = (time.time() - start_time)
print("Accuracy: %s" % acc_dt)
print("Accuracy CV 10-Fold: %s" % acc_cv_dt)
print("Running Time: %s" % datetime.timedelta(seconds=dt_time))
```

```
Accuracy: 100.0
Accuracy CV 10-Fold: 79.59
Running Time: 0:00:00.257872
```

In [ ]:
```
# Gradient Boosting Trees
start_time = time.time()
train_pred_gbt, acc_gbt, acc_cv_gbt = fit_ml_algo(GradientBoostingClassifier(),X_train, y_train,10)
gbt_time = (time.time() - start_time)
print("Accuracy: %s" % acc_gbt)
print("Accuracy CV 10-Fold: %s" % acc_cv_gbt)
print("Running Time: %s" % datetime.timedelta(seconds=gbt_time))
```

```
Accuracy: 93.2
Accuracy CV 10-Fold: 87.17
Running Time: 0:00:02.839144
```

In [ ]:
```
# Random Forest
start_time = time.time()
train_pred_dt, acc_rf, acc_cv_rf = fit_ml_algo(RandomForestClassifier(n_estimators=100),X_train, y_train,10)
rf_time = (time.time() - start_time)
print("Accuracy: %s" % acc_rf)
print("Accuracy CV 10-Fold: %s" % acc_cv_rf)
print("Running Time: %s" % datetime.timedelta(seconds=rf_time))
```

```
Accuracy: 100.0
Accuracy CV 10-Fold: 85.81
```

8. TESTING:

We usually write two different classes of tests for Machine Learning systems:

- Pre-train tests
- Post-train tests

Pre-train tests: The intention is to write such tests which can be run without trained parameters so that we can catch implementation errors early on. This helps in avoiding the extra time and effort spent in a wasted training job.

We can test the following in the pre-train test:

- the model predicted output shape is proper or not
- test dataset leakage i.e. checking whether the data in training and testing datasets have no duplication
- temporal data leakage which involves checking whether the dependencies between training and test data do not lead to unrealistic situations in the time domain like training on a future data point and testing on a past data point
- check for the output ranges. In the cases where we are predicting outputs in a certain range (for example when predicting probabilities), we need to ensure the final prediction is not outside the expected range of values.
- Ensuring a gradient step training on a batch of data leads to a decrease in the loss
- data profiling assertions

Post-train tests: Post-train tests are aimed at testing the model's behavior. We want to test the learned logic and it could be tested on the following points and more:

- invariance tests which involve testing the model by tweaking only one feature in a data point and checking for consistency in model predictions. For example, if we are working with a loan prediction dataset then change in sex should not affect an individual's eligibility for the loan given all other features are the same or in the case of titanic survivor probability prediction data, change in the passenger's name should not affect their chances of survival.
- Directional expectations wherein we test for a direct relation between feature values and predictions. For example, in the case of a loan prediction problem, having a higher credit score should definitely increase a person's eligibility for a loan.

9.ADVANTAGES & DISADVANTAGES:

9.1 ADVANTAGES:

1. Automation of Everything
2. Wide range of applications
3. Scope of improvement

9.2 DISADVANTAGES:

1. High error possibility
2. Data acquisition
3. Low performance of algorithm

It is important to know the advantages and disadvantages of Machine Learning because it will help you. In ways like algorithm designing, decision making, etc. These were some of the most important advantages and disadvantages of Machine Learning.

10. CONCLUSION:

This research addresses the Naive Bayes, KNN and SVM algorithm for Classifying and predicting, whether the employee will leave the company or not. The overall performance of each SVM, KNN and Naive Bayes Algorithms has in contrast in phrases of accuracy. Naive Bayes Algorithm predicted the employee turnover with the accuracy of 76%, KNN with the accuracy of 94% and SVM with the accuracy of 96%. This task concludes, overall performance of SVM algorithm is greater correct over accuracy of Naive Bayesian algorithm and KNN. This SVM mannequin can probably assist a business enterprise to predict the departure of key software program builders and they can be retained in the company, via taking proactive motion such as offering earnings hikes or bendy timing or through higher managing workload variance amongst mission contributors etc. to keep away from massive loss to company. The future enhancement might include some more data science algorithms and deep learning algorithms. Along with Naïve Bayes, KNN and Support Vector Machine, the dataset can be examined with Deep Neural Network and overall performance of the algorithm is evaluated and as soon as the best algorithm has been found, that algorithm will be used to predict whether or not the precise worker will stay or go away the company.

11. FUTURE SCOPE:

Retaining the high-quality personnel ensures patron satisfaction. Increased revenues and comfortable colleagues and staff (Xiao-Li Qu et all., 2015). To hold employees, organizations spend a lot of cash on training, presenting onsite opportunities, and paying above-market salaries. Build a data-driven turnover (Thomas Hugh Feeley et all., 2019) prediction model to predict future turnover (Zubin R et all., 2013), each at combination degrees as properly as for figuring out people with excessive threat of attrition. Using data science classification algorithm, employee turnover can be predicted accurately. Data science is an aggregate of records and AI. Data Science is the series of historic and latest trends in statistics, AI, and machine learning. Data science is a process to do analysis such as frequent pattern mining and association rule mining, classification, clustering and prediction to find the intelligent pattern hidden in the huge amount of dataset. This research predicts whether developers will leave the company after a positive period of time, based on monthly reports. For prediction, Naive Bayes algorithms, KNN and SVM computing device algorithms that used to classify and predict the developer turnover the usage of factors, such as pleasure level, remaining evaluation, wide variety projects, average_monthly_hours, time_spent_company, etc. The overall performance of all the algorithms has in contrast in phrases of accuracy. then the great mannequin will predict the worker turnover. The insights, alongside with data-driven predictive models (Mohammad Nayeem Hasan et all., 2019), can be used to sketch positive

plans for lowering attrition, enhancing retention, lowering attrition expenses and mitigating attrition effects.

12. APPENDIX:

Source link:

Sprint 1: https://github.com/IBM-EPBL/IBM-Project-29038-1660120313/tree/main/Project%20Development%20Phase/Sprint%201

Sprint 2: https://github.com/IBM-EPBL/IBM-Project-29038-1660120313/tree/main/Project%20Development%20Phase/Sprint%202

Sprint 3: https://github.com/IBM-EPBL/IBM-Project-29038-1660120313/tree/main/Project%20Development%20Phase/Sprint%203

Sprint 4: https://github.com/IBM-EPBL/IBM-Project-29038-1660120313/tree/main/Project%20Development%20Phase/Sprint%204

TEAM MEMBERS:

Jeyaram K

Aishwarya Lakshmi T

Jagathesh Kumar S

Rajaraman K