**ABSTRACT**

This project is a modular application of HR system software. As a HR system deals with all the tasks handled by a manual HR of a company such as leave management, HR metrics, Recruitment, Employee Relations, Satisfaction, Communication, Retention, Performance Management . Similarly my project is a smaller scale instance of the same where it handles one part of the whole system which is performance review but instead helps by predicting the review beforehand. It predicts the attrition and retention rate of any company making the whole system more efficient as it's an automatic software implementation instead of humans governing everything manually

The entire program has been developed in Python and uses the Eclipse IDE for running the python application.

The mini-project is completely based on the high-level language, Python and uses GUI programming to provide a simple and easy to understand platform for the users.

**CHAPTER 1**

**INTRODUCTION**

**1.1 PROBLEM DEFINITION**

This project is an instance of HR system software with modular implementation of performance evaluation system modified according to specific requirements of the company. This project will be customized according to the needs of a company as many such software exists for a hefty price but they're not customized according to the requirements of the company. What makes this project better is implementation of machine learning on the idea. This will help us predict the performance of an employee better

**1.1.1 OBJECTIVES**

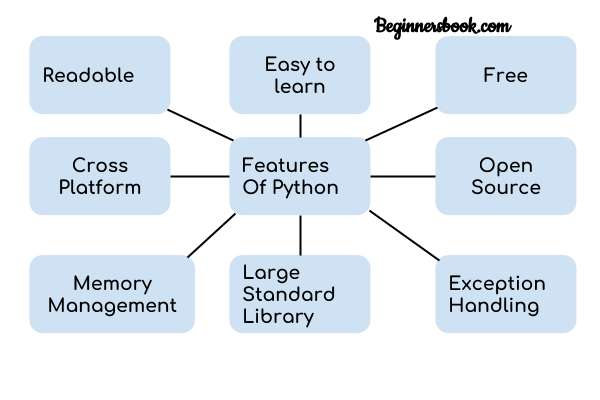
Once developed, the application will provide services to companies which will help manual HR to take some load off as they wont have to manually rate each and every employee but the program will automatically do it for them.

* More efficient implementation
* Works as appraisal system
* Helps improve workforce efficiency by analyzing weaknesses
* Predicts attrition and retention rate
* Improves outcome of company
* Helps in more value for money per employee

**CHAPTER 2**

**FUNDAMENTALS OF PYTHON**

**2.1 INTRODUCTION TO PYTHON**

Python is a regularly and widely utilized broadly useful, significant level programming language. Guido van Rossum in 1991 was the author of Python and was subsequently evolved by Python Software Foundation. It was basically intended to underscore on code meaningfulness, and its linguistic structure permits software engineers to communicate thoughts in couple of lines of code. Python is a user friendly language because :

**Fig 2.1: features of python**

**Python applications:**

1. Web improvement - Web system like Django and Flask depend on Python. They assist you with composing server side code which assists you with overseeing information base, compose backend programming rationale, planning urls and so forth

2. AI - There are many AI applications written in Python. AI is a method for composing a rationale so a machine can learn and take care of a specific issue all alone. For instance, items suggestion in sites like Amazon, Flipkart, eBay and so forth is an AI calculation that recognises client's advantage. Face acknowledgment and Voice acknowledgment in your telephone is one more illustration of AI.

3. Information Analysis - Data investigation and information perception in type of graphs can likewise be created utilizing Python.

4. Prearranging - Scripting is composing little projects to robotize basic errands, for example, sending computerized reaction messages and so forth Such kind of utilizations can likewise be written in Python programming language.

5. Game turn of events - You can foster games utilizing Python.

6. You can foster Embedded applications in Python.

7. Work area applications - You can foster work area application in Python utilizing library like TKinter or QT.

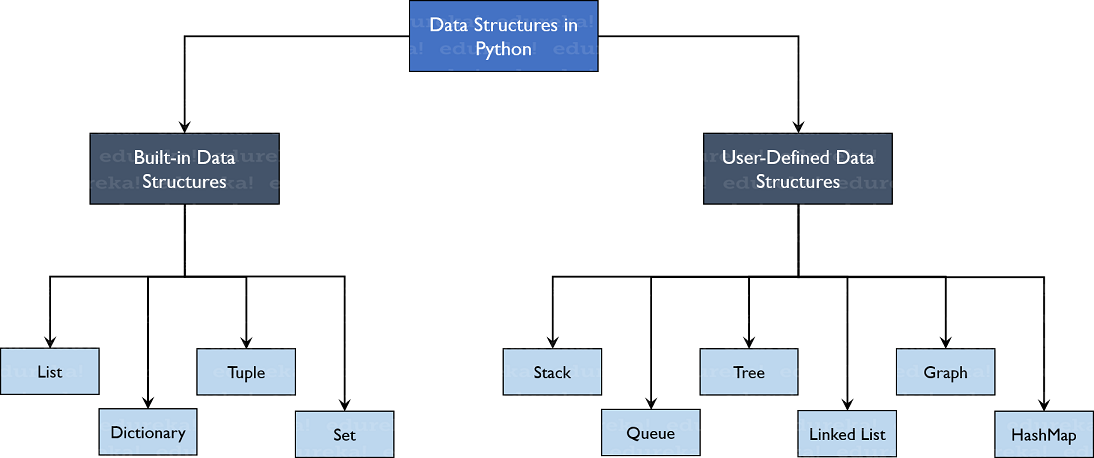
**What is a Data Structure?**

Sorting out, overseeing and putting away information is significant as it empowers more straightforward access and proficient changes. Information Structures permits you to put together your information so that empowers you to store assortments of information, relate them and perform procedure on them likewise.

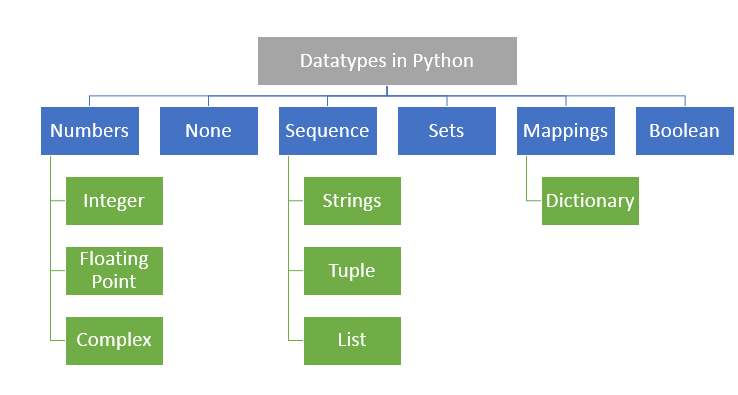
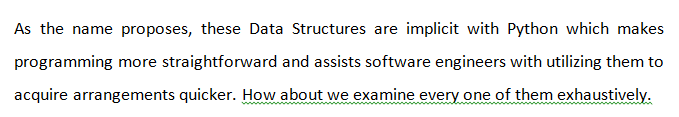
**Kinds of Data Structures in Python**

Python has certain help for Data Structures which empower you to store and access information. These designs are called List, Dictionary, Tuple and Set.

Python permits its clients to make their own Data Structures empowering them to have full command over their usefulness. The most conspicuous Data Structures are Stack, Queue, Tree, Linked List, etc which are additionally accessible to you in other programming dialects. So since you realize what are the sorts accessible to you, for what reason don't we push forward to the Data Structures and carry out them utilizing python.

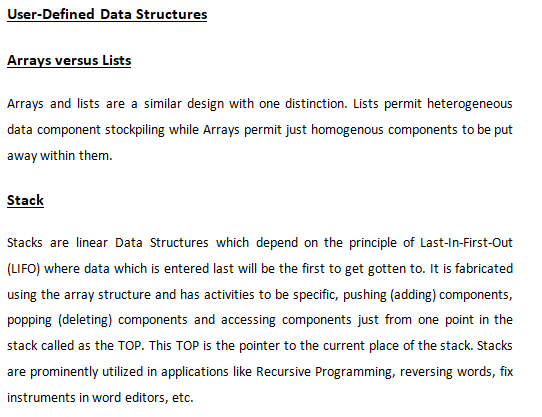
**Built in Data Structures**

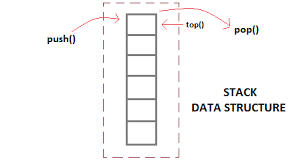
**Fig 2.2: DS in python**



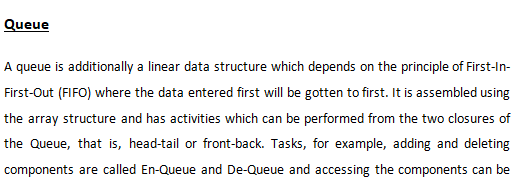
**Fig 2.3: data types in python**

-

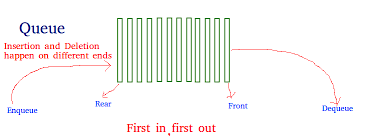




**Fig 2.4: stack**

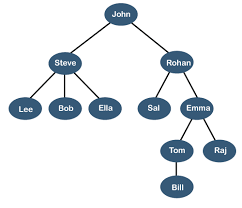
****

performed. Queues are utilized as Network Buffers for gridlock the executives, utilized in Operating Systems for Job Scheduling and some more.

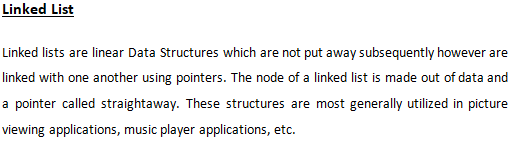


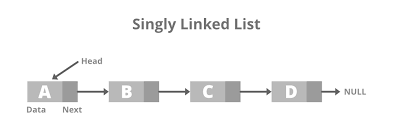
**Fig 2.5: queue**

**Tree**

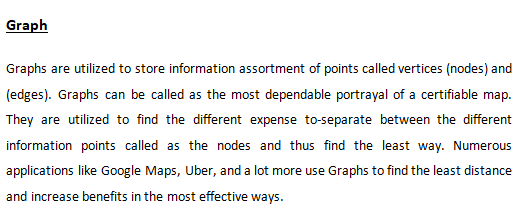
Trees are non-linear Data Structures which have root and nodes. The root is the node from where the data originates and the nodes are different data points that are accessible to us. The node that goes before is the parent and the node after is known as the youngster. There are levels a tree needs to show the profundity of information. The last nodes are known as the leaves. Trees make a progressive system which can be utilized in a ton of true applications, for example, the HTML pages use trees to distinguish which label goes under which block. It is additionally productive in searching purposes and significantly more.

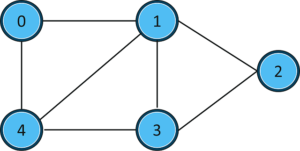
**Fig 2.6: tree**

****

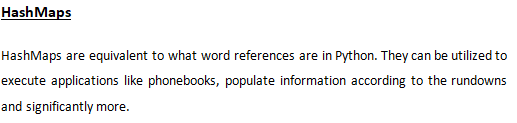


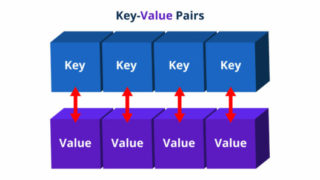
**Fig 2.7: linked list**

****



**Fig 2.8: graphs**

****



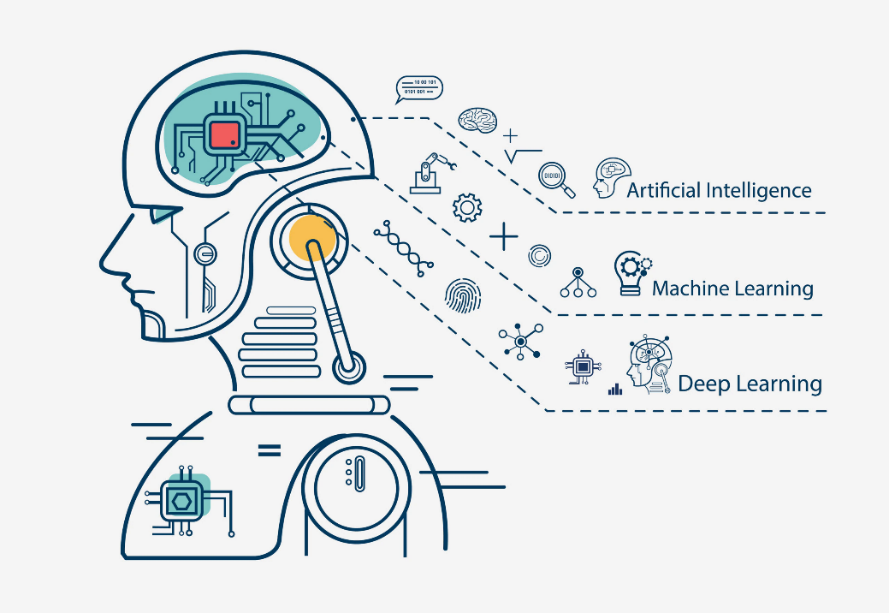
**Fig 2.9: HashMaps**

**CHAPTER 3**

**3.1 Machine learning:**

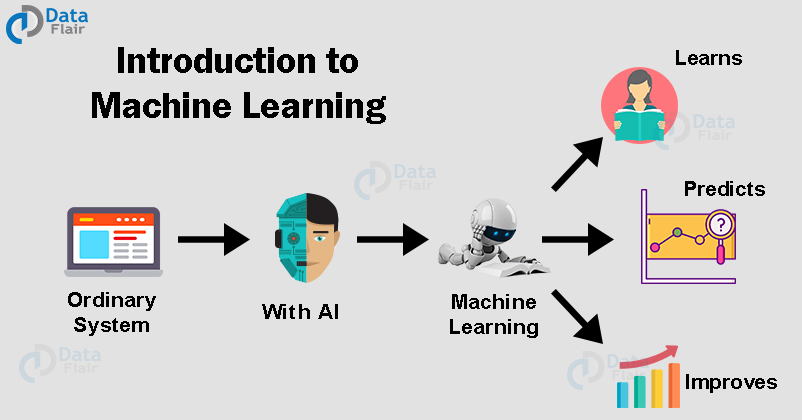
Machine learning is a piece of computerized reasoning (AI) and PC programming which bases on the utilization of information and calculations to mirror the way that people learn, reliably working on its precision.

AI is the field of study that gives PCs the ability to learn without being expressly customized. ML is one of the most thrilling advances that one would have at any point run over. As it is obvious from the name, it gives the PC that makes it more like people: The capacity to learn. AI is effectively being utilized today, maybe in a lot a bigger number of spots than one would anticipate.

Machine Learning(ML) can be clarified as mechanizing and further developing the learning system of PCs in light of their encounters without being really customized for example with no human help. The cycle begins with taking care of good quality information and afterward preparing our machines(computers) by building AI models utilizing the information and various calculations. The selection of calculations relies upon what sort of information do we have and what sort of assignment we are attempting to robotize.

**Fig 3.1: Machine learning**

ML is a critical piece of the making field of information science. Using obvious techniques, calculations are prepared to make arrangements or suspicions, uncovering key snippets of data inside information mining projects. These snippets of data similarly drive choice creation inside applications and affiliations, preferably affecting key improvement assessments. As colossal information proceeds to widen and make, the market income for information researchers will increment, guessing that they should help the recognizable proof of the most material business questions and accordingly the information to respond to them.



**Fig 3.2: introduction to ML**

Example of how ML work

We use SIRI to ask questions like “hi Siri how far is the nearest Subway”, a powerful speech recognition kicks off and converts the audio into its corresponding textual form which is then sent to the Apple service for the processing then neural network processing algorithms are run to understand the user's intent and finally Siri tells the answer. well this is what machine learning is all about making machines learn and act like humans by feeding them with data and information without been explicitly programmed as we saw in the previous example when the data counting machines immediately start analysing data and eventually gets trained on it and learn it now when you date a point comes in machine accurately makes prediction and decisions based on the past data now that you know what is machine learning let's talk about supervised and unsupervised learning.

**Supervised learning**

Supervised learning as the name suggests works under supervision that is its learning in which machine is trained data which is well labelled and then predicts with the help of the labelled data but what is a labelled data set? date for which you already know the target answer is called a labelled data. let me show you an image and tell you that it's a dog then its a labelled data while if I show you an image without telling you what exactly it is then it's an unlabeled data. we have images which are labelled as spoon and knife. We then feed it to the machine which analyses and learns the association of these images with its labels based on its features Such as shape, size, sharpness etc. now when a new images fed the machine without any label with the help of the past and the machine is able to predict accurately and tell that it was a spoon and hence supervised machine learning that teaches the model to learn from the label example that we provide.

So supervised learning can be further divided into : **classifications and regression.**

It is a classification problem when the output variable is categorical such as red or blue disease or no disease male female there as its regression model when the output variable is it real or continuous for example salary based work experience weight based on height should create a predictive model showing trends and data.

**Example 1**: If I say will I get a salary raise or not its classification and if I ask how much salary raise will I get that's regression. But if I say how much salary raise will I get that is regression.

Now let understand **classification** to help of an example. In order to predict whether an email is a spam or not, first we teacher machine what is it really looks like. This is done based on a lot of spam filters leg was reviewing the content of the email then review the email header and search which contains any falsified information based on some keywords like free lottery, prize claim it then gently Blacklist filters to stop emails at come from already blacklisted known spammers. So all these filters scores the email which is known as spam school the lower the total spam score of the email it is more likely that email will land in subscribers inbox so based on the continent labels and spam score of the new incoming name the algorithm decides whether it should be managed in inbox of the spam folder. Now lets quickly understand **regression** analysis. We have two variables that the temperature and humidity and temperature, temperature is the independent variable and humidity is the dependent variable such that as temperature increases humidity decreases in the correlated when we stayed at our discretion model it will understand relationship between these two variables and how one variable depends on the other after the machine is trained it can easily project to the humidity based on the given temperature. Well that was about regression.

Now let's see some real life applications for supervised learning is used for supervised learning is used in risk assessment to assess risk in financial services or an insurance to to minimise the risk portfolio of the company's. Image classification Facebook recognises your friend in a picture from it and move that photos image classification is one of the keys cases of demonstrating supervised machine learning algorithms of sleep a lot more goes into all these like convolutional neural networks for fraud detection whether the transactions made by the user automatic or not and visual recognition the ability of a machine learning model to identify objects places people in actions in images.

**Unsupervised learning**

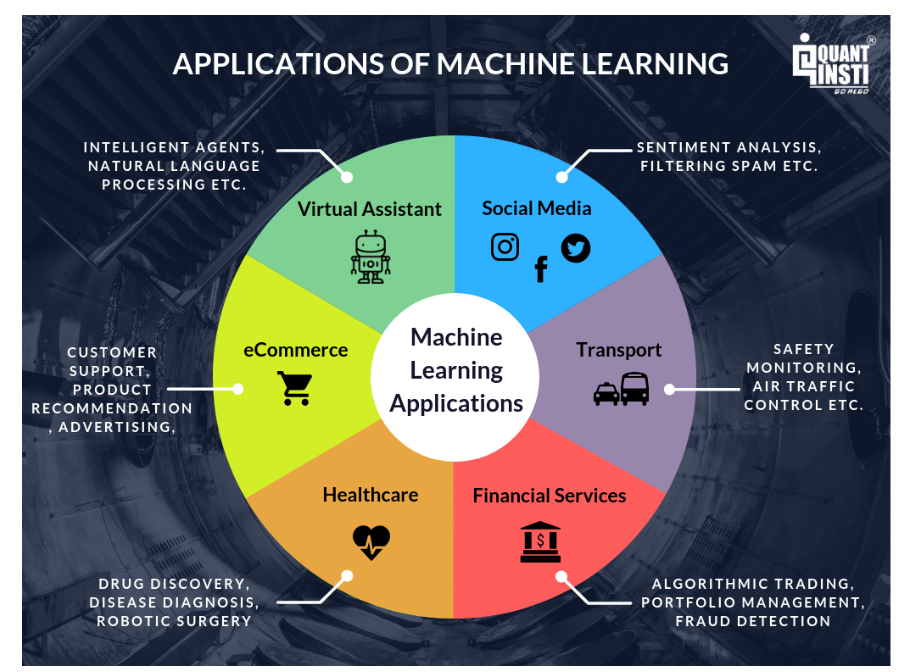
In unsupervised learning there is no super mention that is no training will be given to the machine allowing it to act on the data which is not labelled hence machine tries to identify patterns and get the response take a similar example as before this time we do not tell the machine whether it's Pune and if the machine identifies patterns for the given set and groups them based on the patterns similarities again and supervised learning can be for the grouped into clustering and association clustering is basically by the machine forms groups based on the behaviour of the data secondly association it is a rule-based machine learning to discover interesting relation between variables in large data sets for example which customer made similar product purchases is clustering there as association is which products were purchased together now let understand clustering with the help of an example to reduce the churn rate at telecom companies studies the behaviour of the customers based on average quality ration and internet usage and observes that while some customers call duration is quite high others have heavy internet usage the customers are grouped based on the observed behaviour analyst strategies adopted to minimise generate and maximize profit by suitable promotions campaigns as you can see the chart and the right inside customers in group is more data and also have high quality ration Ruby customers are heavy internet users by group c customers have high quality ration so group b will be given more data benefit plans by group c will be given cheaper call rates to buy their loyalty so this was the example of clustering now let understand association with another example let's say one goes to supermarket and buy this product se bread milk fruits feet and customer to NGOs and wise bread milk rice and better now when customer 3 goes in b gread it is highly likely that you will also find it hence relationship is established based on customer behaviour and recommendations Amit now let's look at some real life applications of unsupervised learning market basket analysis in machine learning model based on the algorithm that if you buy a certain group of items you are less more likely to buy another group of items cement plastering cement similar words share similar contacts people post queries on website On based clustering groups all responses in a cluster with same meaning to ensure that the customer find the information they want quickly and easily it plays an important role in information retrieval good browsing experience in comprehension delivery store optimisation machine learning models are used to predict the demand and keep up with the supply also to open stores when demand is more and optimising rules for more efficient delivery going to pass data and behaviour can also use unsupervised machine learning models to identify accident prone areas based on the intensity of those accidents and the area in order to introduce safety measures by now I hope you want to supervised and unsupervised learning for a week and lets you are few differences between the two the most fundamental difference is that supervised learning uses known as legal data and unsupervised learning uses and label it has the input secondary supervised learning follows feedback mechanism and unsupervised learning does not also the most commonly used algorithms in supervised learning at decision tree logistic regression support vector machine accept and unsupervised learning k means clustering hierarchical clustering algorithm.

**3. Reinforcement Learning:**

How it functions: Using this algo, the machine is prepared to settle on explicit choices. It works thusly: the machine is presented to a climate where it trains itself constantly utilizing experimentation. This machine gains from previous experience and attempts to catch the most ideal information to settle on precise business choices. Illustration of Reinforcement Learning: Markov Decision Process

Working of Machine Learning processes:

1. Decision Process: by and large, AI calculations are utilized to make a figure or characterization. Considering a few information, which can be named or unlabeled, your calculation will make a check about a model in the information.
2. An Error Function: A mistake work serves to study the figure of the model. In the event that there are known models, a screw up cutoff can make a relationship with survey the precision of the model.
3. A Model Optimization Process: If the model fits better to the server farms in the arranging set, then, at that point, loads are acclimated to lessen the goof between the known model and the model check. The calculation will go over this assess and further foster measure, resuscitating troubles freely until an edge of precision has been met.develop measure, reviving burdens independently until an edge of exactness has been met.

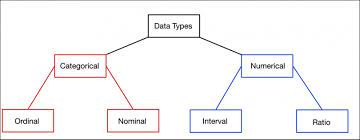


**Fig 3.3: ML applications**

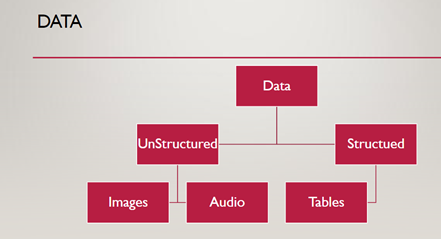
**Data and collection of data**

Data are facts or information usually numeric that are collected through to making observations.

It is set of values of quantitative or qualitative variables about one or more persons objects used for discussion, calculation or reasoning



**Fig 3.4: DATA TYPES**



**Fig 3.5: TYPES OF DATA**

**Data**

Qualitative/Categorical

* Nominal
* Ordinal

Quantitative / Numerical

* Discrete
* Continuous

**BASICS STEPS OF ML MODEL**

1. Collect Data

2. Train Model

• Iterate many times until it is good enough

3. Deploy Model

Get data back and remodel

**DATA IS MEESSY**

1. Garbage in, garbage out
2. Data Problems:

• Incorrect Values

* Missing values

**EXPLORATORY DATA ANALYSIS**

•In this stage, information engineers have a few inquiries close by and attempt to approve those inquiries by performing EDA.

•Exploratory Data Analysis, or EDA, is basically a kind of narrating for analysts.

•It permits us to reveal examples and experiences, frequently with visual techniques, inside information. EDA is frequently the initial step of the information demonstrating process

•Nonetheless, it's not really truly challenging to play out an EDA.

•EDA might sound colorful on the off chance that you are new to the universe of measurements.

**EVALUATING REGRESSION MODELS**

1.Mean Absolute Error

2.Mean Squared Error

3.R-Squared

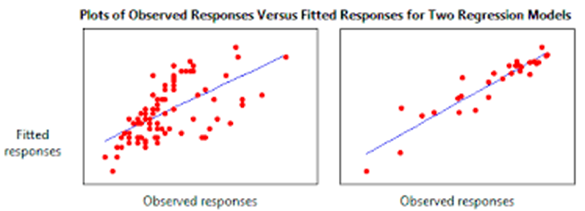
4.Adjusted R-Squared

1. R-SQUARED METHOD

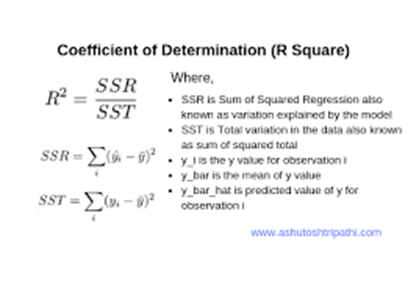
R Squared is an estimation that lets you to know degree the extent of difference in the reliant variable clarified by the fluctuation in the autonomous factors. In more straightforward terms, while the coefficients gauge patterns, R-squared addresses the spread around the line of best fit.

•For instance, in the event that the R² is 0.80, 80% of the variety can be clarified by the model's bits of feedbacks.

•Assuming the R² is 1.0 or 100 percent, that implies that all developments of the reliant variable can be totally clarified by the developments of the free factors.



**Fig 3.6: comparision of model with low and high**



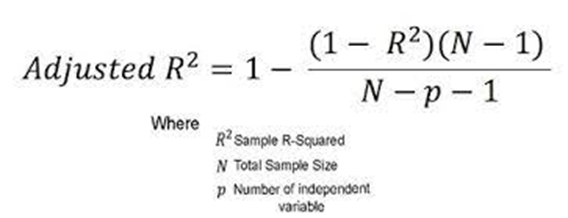
**Fig 3.7: r squared formula**

1. ADJUSTED R-SQUARED METHOD

Each extra autonomous variable added to a model generally builds the R² esteem - accordingly, a model with a few free factors might appear to be a superior fit regardless of whether it isn't.

•This is the place where Adjusted R² comes in.

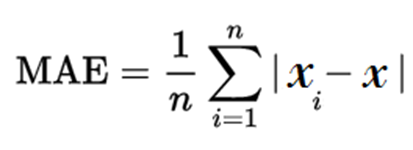
•The changed R² makes up for each extra autonomous variable and possibly increments in the event that each given variable works on the model above what is conceivable by likelihood.



**Fig 3.8: adjusted r square formula**

1. MEAN ABSOLUTE ERROR(MAE)

The absolute errorr is the contrast between the actual and predicted val. Consequently, the MAE is the normal of the absolute error.

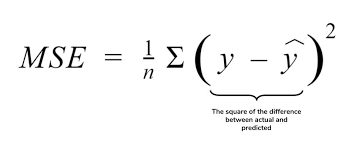


**Fig 3.9: MAE formula**

1. MEAN SQUARED ERROR(MSE)

The mean squared errorr or MSE is like the MAE, with the exception of you take the normal of the squared contrasts between the actual and predicted val.

Since the distinctions are squared, bigger mistakes are weighted all the more exceptionally, thus this ought to be utilized over the MAE when you need to limit large errors. The following is the condition for MSE, just as the code.



**Fig 3.10: MSE formula**

**ASSESSING CLASSIFICATION MODELS**

1.AUC-ROC Curve

2.Confusion Matrix and related metrics

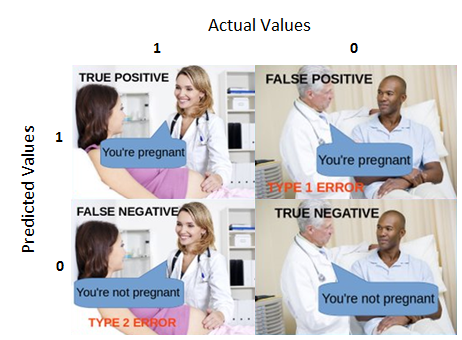
3.F1 Score

**CONFUSION MATRIX:**

It is a matrix with 2 major classification i.e, positive and negative and further division true and false so : TP, FP, TN, FN

It is very valuable for estimating Accuracy, Specificity, Precision, Recall and in particular AUC-ROC bends.

**Fig 3.11: confusion matrix**

Lets understand using an example:

**Fig 3.12: example to understand confusion matrix**

* True Positive:

Understanding: we calculated output to be positive and the result was same(T).

* True Negative:

Understanding: we calculated output to be negative and the result was not same(T).

.False Positive: (Type 1 Error)

Understanding: we calculated output to be positive and the result was not same (F).

* False Negative: (Type 2 Error)

Translation: we calculated output to be negative and the result was same (F).

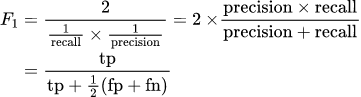
Remember,

Positive = true

Negative = false

**F-score**

The F1 score is a proportion of a test's exactness - it is the harmonic mean of recall & precision. It can attain a greatest score of 1 (perfect recall & precision) and at least 0. Generally, it is a proportion of the accuracy and vigor of your model.



**Fig 3.13: f score**

**Instances of machine learning applications:**

Speech Recognition: It is a limit which uses standard language taking care of (NLP) to manage human talk into a created design. Various cells solidify talk affirmation into their structures to lead voice look for instance Siri-or give more noteworthy accessibility around informing.

Client care: Online chatbots are superseding human experts along the customer adventure. They answer sometimes presented requests around subjects, like transportation, or give modified appeal, decisively pitching things or proposing sizes for customers, changing the way wherein we consider customer responsibility across destinations and online media stages. Models informed bots for online business objections with virtual subject matter experts, illuminating applications, similar to Slack and Facebook Messenger, and tasks regularly done by modest partners and voice partners.

Mechanized stock trading: AI-driven high-repeat , Designed to smooth out stock portfolios, trading stages make thousands or even extraordinary many trades every day without human intervention.

Here is the rundown of generally utilized AI algos. These calculations practically used for:

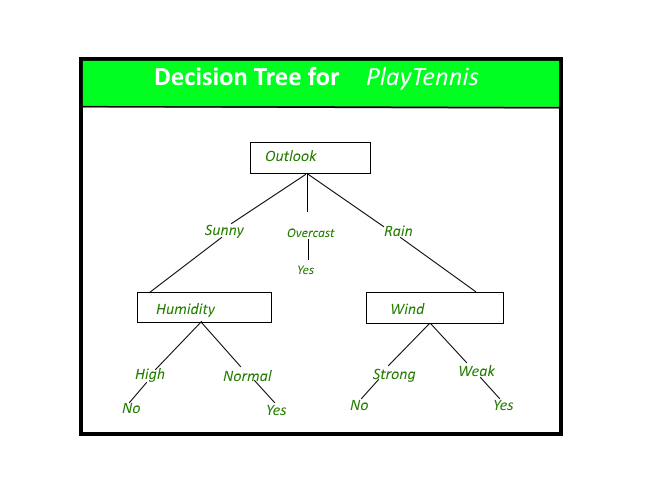
* Gradient Boosting algorithms
  1. XGBoost
  2. CatBoost
  3. GBM
  4. LightGBM
* Random Forest
* Decision Tree
* SVM
* Naive Bayes
* Linear Regression
* K-Means
* Logistic Regression
* Dimensionality Reduction Algorithms
* kNN

**3.2** **Decision Tree:**

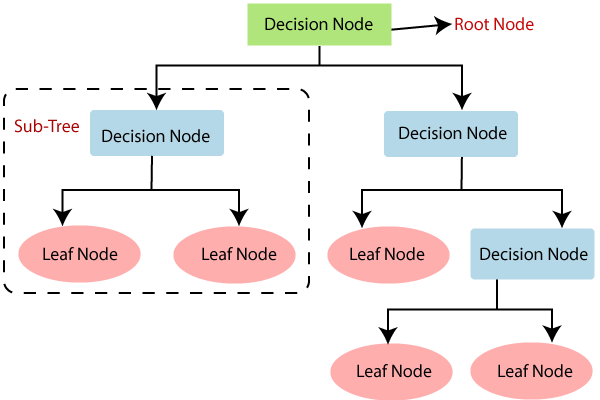
Decision tree is highly noteworthy aslo notable instrument for forecast and classification. Its is a flowchart like tree structure, where each internal node demonstrates a test on a property, each branch tends to a consequence of the test, and each leaf node holds a class mark.

If all else fails, decision trees are managed an algorithmic way of thinking that recognizes ways of managing segment an informational record subject to various conditions. It is perhaps the most completely utilized and supportive framework for coordinated learning. Choice Trees are a non-parametric composed learning approach utilized for both characterization and relapse issues. The objective is to make a model that predicts the worth of an objective variable by taking in clear choice guidelines got from the information highlights.

The choice standards are all things considered in sort of on the off chance that

clarifications. The more significant the tree, the seriously bewildering the rules and fitter the model.

**Fig 3.7: decision tree**

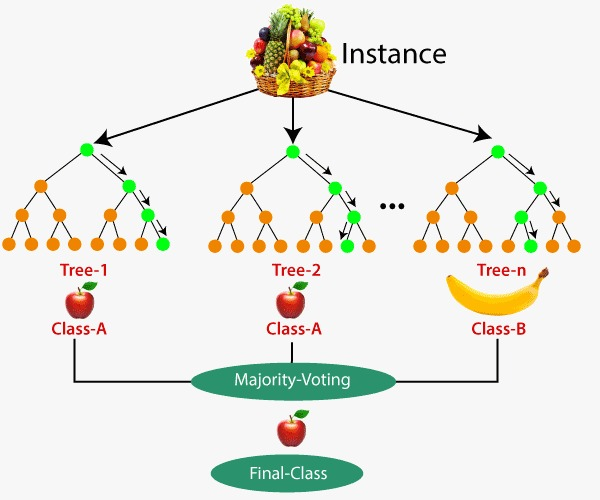


**Fig 3.8: decision node**

* 1. **Random Forest Classifier:**

Random forest is an overseen learning algo. It very well may be utilized both for backslide and classification. It is additionally the most adaptable and simple to utilize assessment. A forest is incorporated trees. It is said that the more trees it has, the heartier a forest is. Random forest make choice trees on arbitrarily picked information tests, gets presumption from each tree and picks the best course of action through projecting a surveying structure. It besides gives an amazingly decent pointer of the part importance.

The underneath chart clarifies the working of the Random Forest calculation :

**Fig 3.9: decision tree** 

**Suppositions for Random Forest**

Since the random forest joins different trees to foresee the class of the dataset, it is conceivable that some choice trees might anticipate the right result, while others may not. Be that as it may, together, every one of the trees foresee the right result. Accordingly, underneath are two presumptions for a superior Random timberland classifier:

There should be a few genuine qualities in the element variable of the dataset with the goal that the classifier can foresee precise outcomes rather than a speculated outcome.

The expectations from each tree should have extremely low relationships.

Why utilize Random Forest?

The following are a few focuses that clarify why we should utilize the Random Forest calculation:

It requires some investment when contrasted with different calculations.

It predicts yield with high exactness, in any event, for the huge dataset it runs productively.

It can likewise keep up with exactness when a huge extent of information is absent.

How does Random Forest calculation function?

Arbitrary Forest works in two-stage initially is to make the irregular woods by consolidating N choice tree, and second is to make expectations for each tree made in the main stage.

The Working system can be clarified in the underneath steps and chart:

Step-1: Select arbitrary K informative elements from the training set.

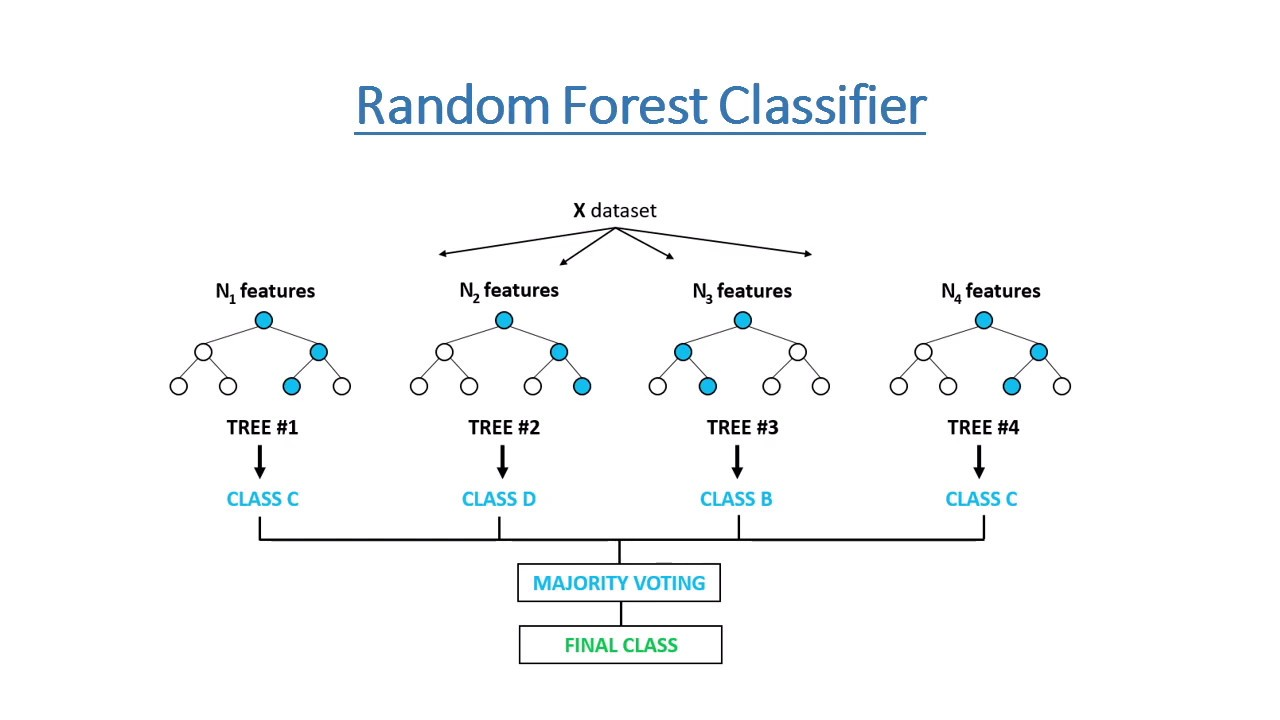
Step-2: Build the decision trees related with the chose elements (Subsets).

Step-3: Choose any number N for decision trees that you need to construct.

Step-4: Repeat Step 1 and 2.

Step-5: For new elements, observe the forecasts of every choice tree, and dole out the new information focuses to the classification that successes the greater part casts a ballot.

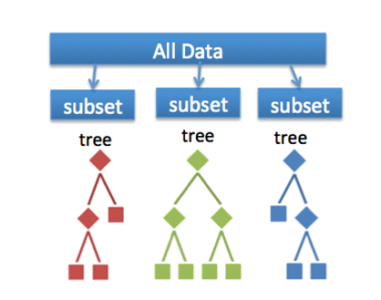
The working of the calculation can be better perceived by the beneath model:

**Fig 3.10: random forest classifier** 

Random forests have a combination of employments, similar to proposition engines, picture classification, and part assurance. It might be used to classify enduring progressed up-and-comers, identify tricky activity and expect ailments. It lies at the establishment of the Boruta computation, which picks significant arrangements in a dataset.

It really is an outfit method (in light of the gap and-overcome approach) of decision trees created on a randomly split dataset. This arrangement of decision tree classifiers is generally called the forest. The singular decision trees are made using a characteristic decision pointer, for instance, information obtain, procure extent, and Gini record for every quality. Each tree depends upon a free random model. In a classification issue, each tree votes, and the most standard class is picked as the finished result. By virtue of

backslide, the ordinary of all the tree yields is considered as the final result. It is more clear and even more exceptional diverged from the other non-direct classification estimations.



**Fig 3.11: creation of trees**

**Utilizations of Random Forest**

Lets talk about fields in which RF is mostly utilized:

Promoting: the patterns of marketing can be classified.

Use of land: we can compare big /huge areas of land.

Medication: using this method we can easily identify pattern of sickness.

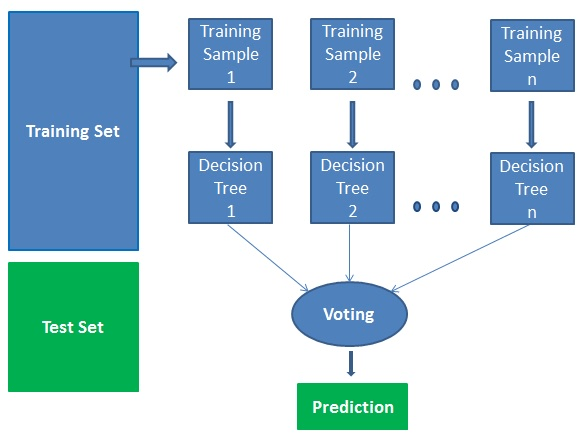
Banking: it is highly used in banking to foresee any credential hazards.

**Impediments of Random Forest**

Albeit random forest can be utilized for both arrangement and relapse assignments, it isn't more reasonable for Regression undertakings.

Python Implementation of RFA

Presently we will use python to execute the RFA tree. Forthis, wewill utilize the equivalent dataset "user\_data.csv", which has been used in past grouping models. By utilizing the equivalent dataset, we can contrast the RF classifier and other characterization models, for example , KNN, Decision tree Classifier, SVM, Logistic Regression, and so forth



**Fig 3.12: training and testing set**

Advantages:

1.Random forest is considered an extraordinarily careful and strong methodology by excellence of how much choice trees taking an interest in the mean time.

2.It doesn't experience the malicious effects of the overfitting issue. The focal explanation is that it takes the conventional of the enormous number of suspicions, which balances propensities.

3.The computation can be used in both arrangement and backslide issues.

4.Random forest can in addition oversee missing qualities. There are two distinct ways to deal with deal with these: utilizing focus qualities to supersede unlimited factors, and taking care of the closeness weighted common of missing attributes.

5.You can get the general part importance, which helps in picking the most contributing plans for the classifier.

**IMPLEMENTATION**

In my project the attributes initially on which data will be evaluated and a prediction would be made are:

EmpNumber,

Age,

Gender,

EducationBackground,

MaritalStatus,

EmpDepartment,

EmpJobRole,

BusinessTravelFrequency,

DistanceFromHome,

EmpEducationLevel,

EmpEnvironmentSatisfaction,

EmpHourlyRate,

EmpJobInvolvement,

EmpJobLevel,

EmpJobSatisfaction,

NumCompaniesWorked,

OverTime,

EmpLastSalaryHikePercent,

EmpRelationshipSatisfaction,

TotalWorkExperienceInYears,

TrainingTimesLastYear,

EmpWorkLifeBalance,

ExperienceYearsAtThisCompany,

ExperienceYearsInCurrentRole,

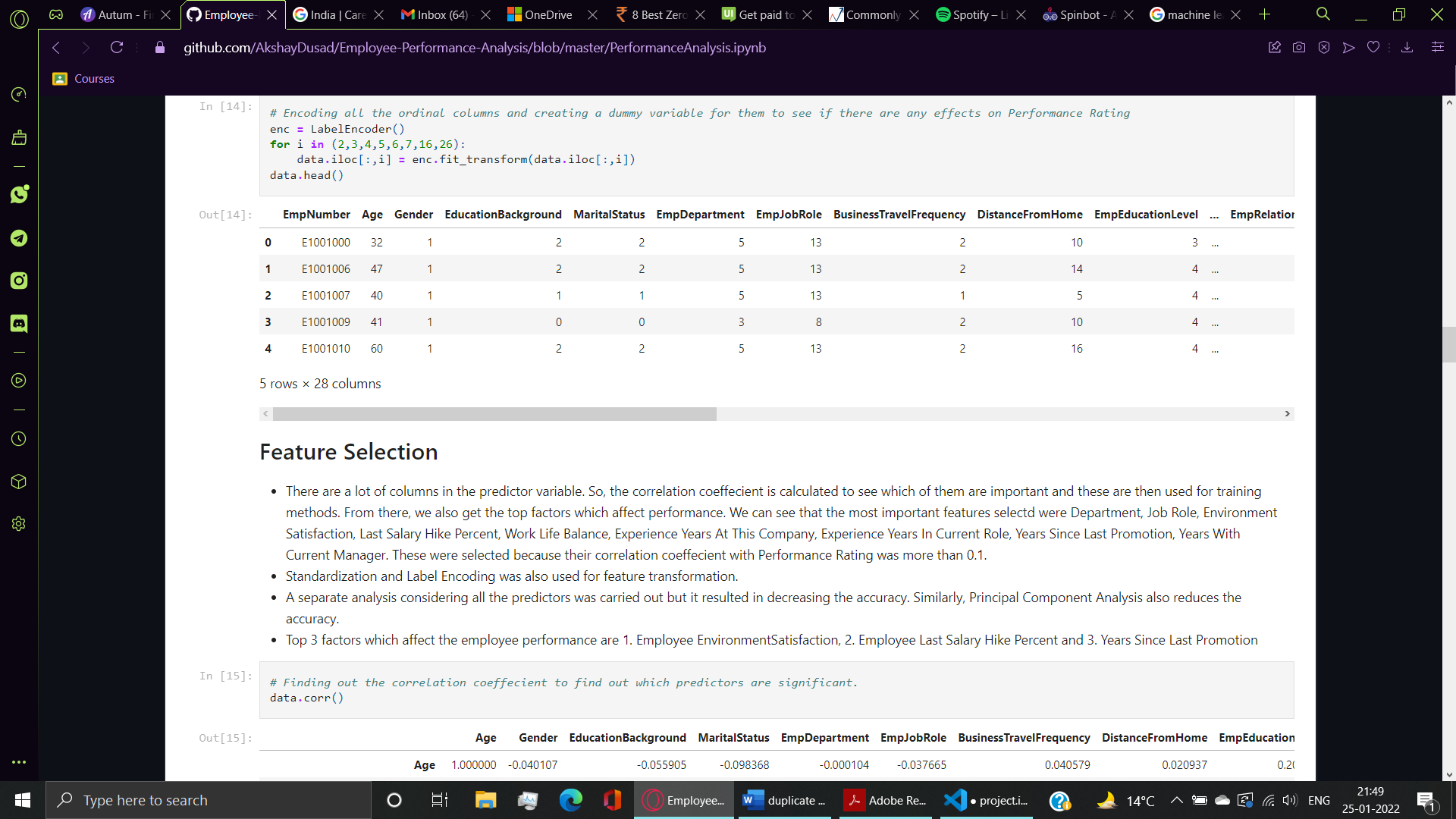
YearsSinceLastPromotion,

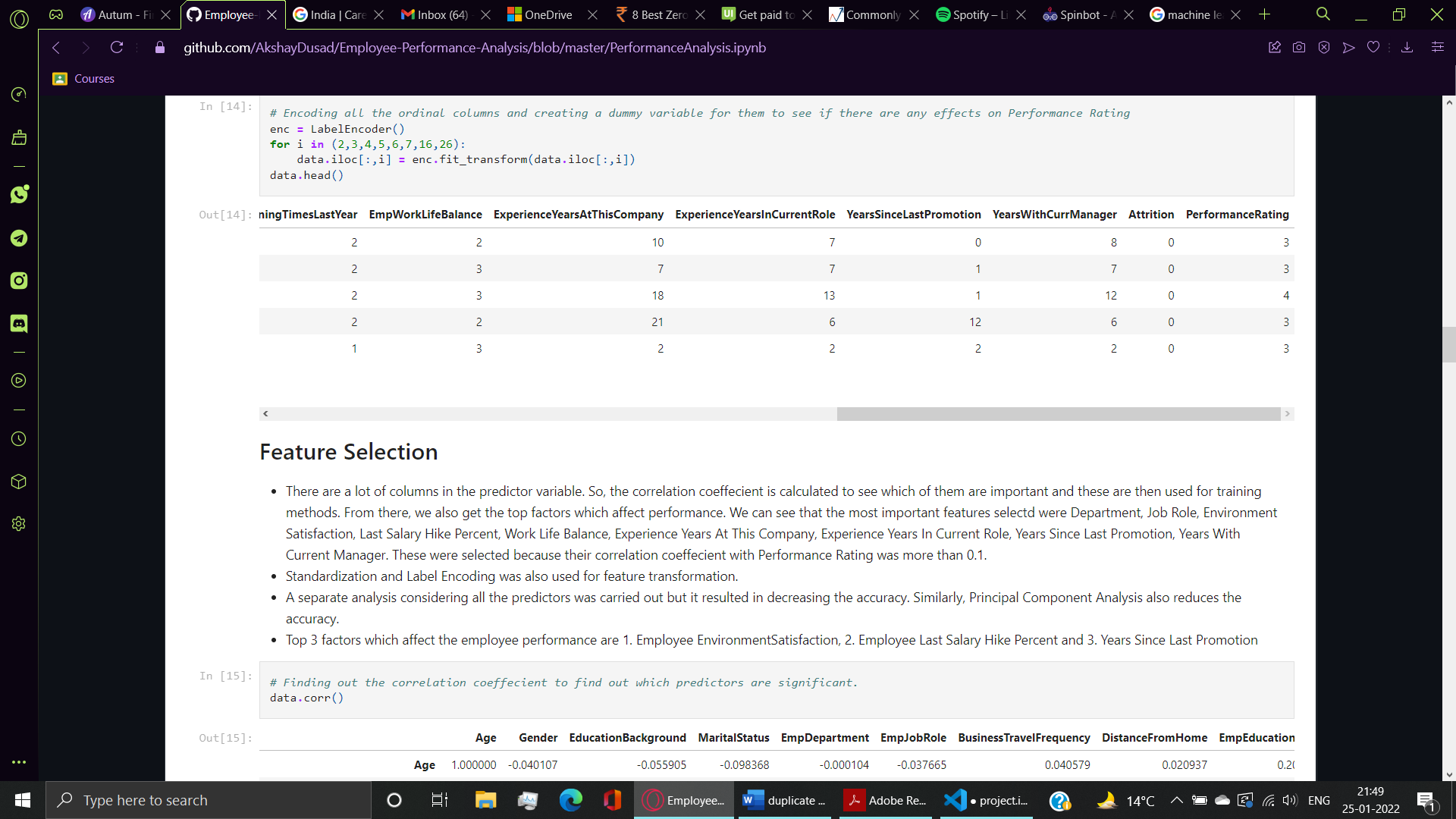
YearsWithCurrManager,

Attrition,

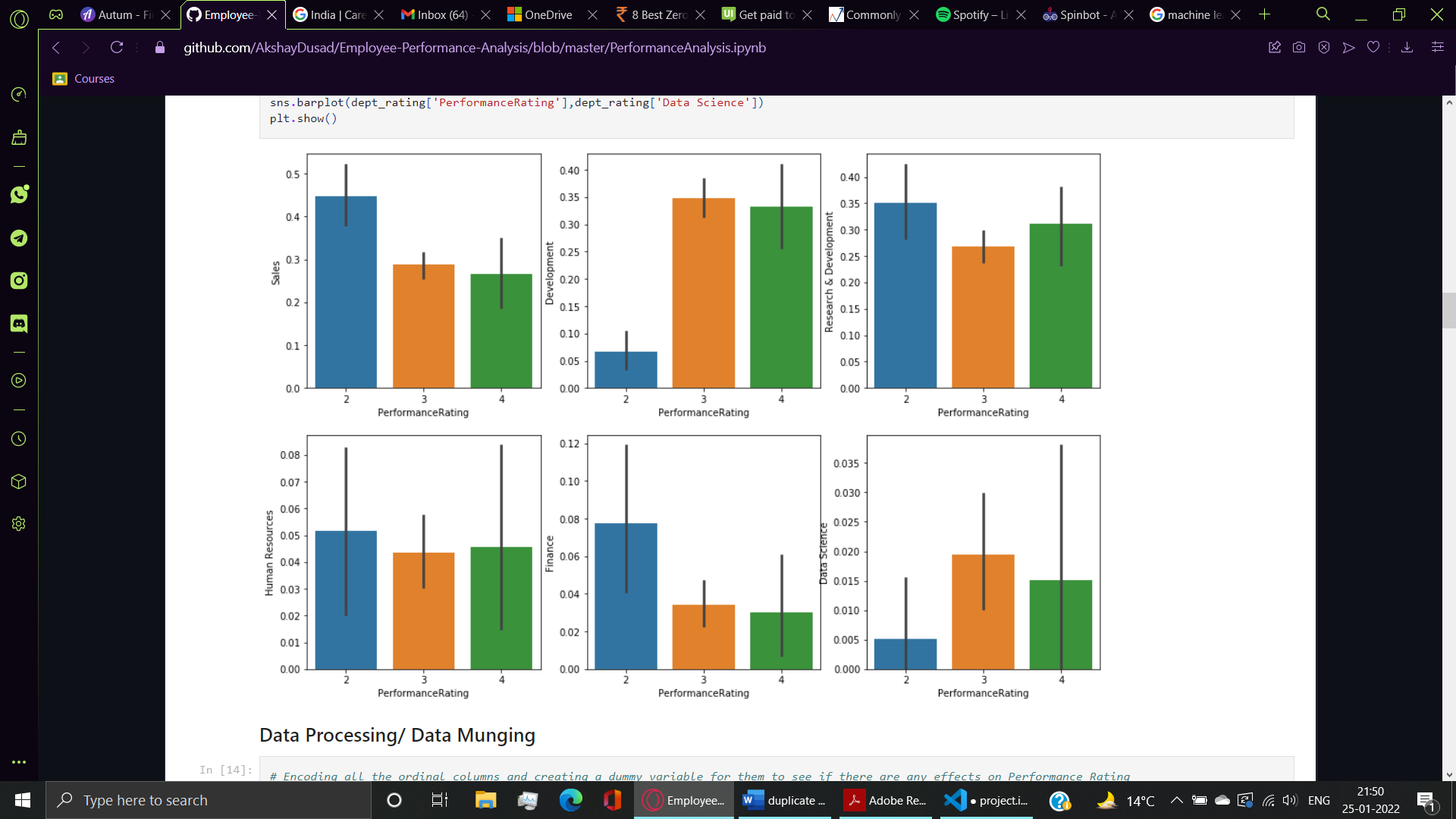
PerformanceRating

How the data looks:

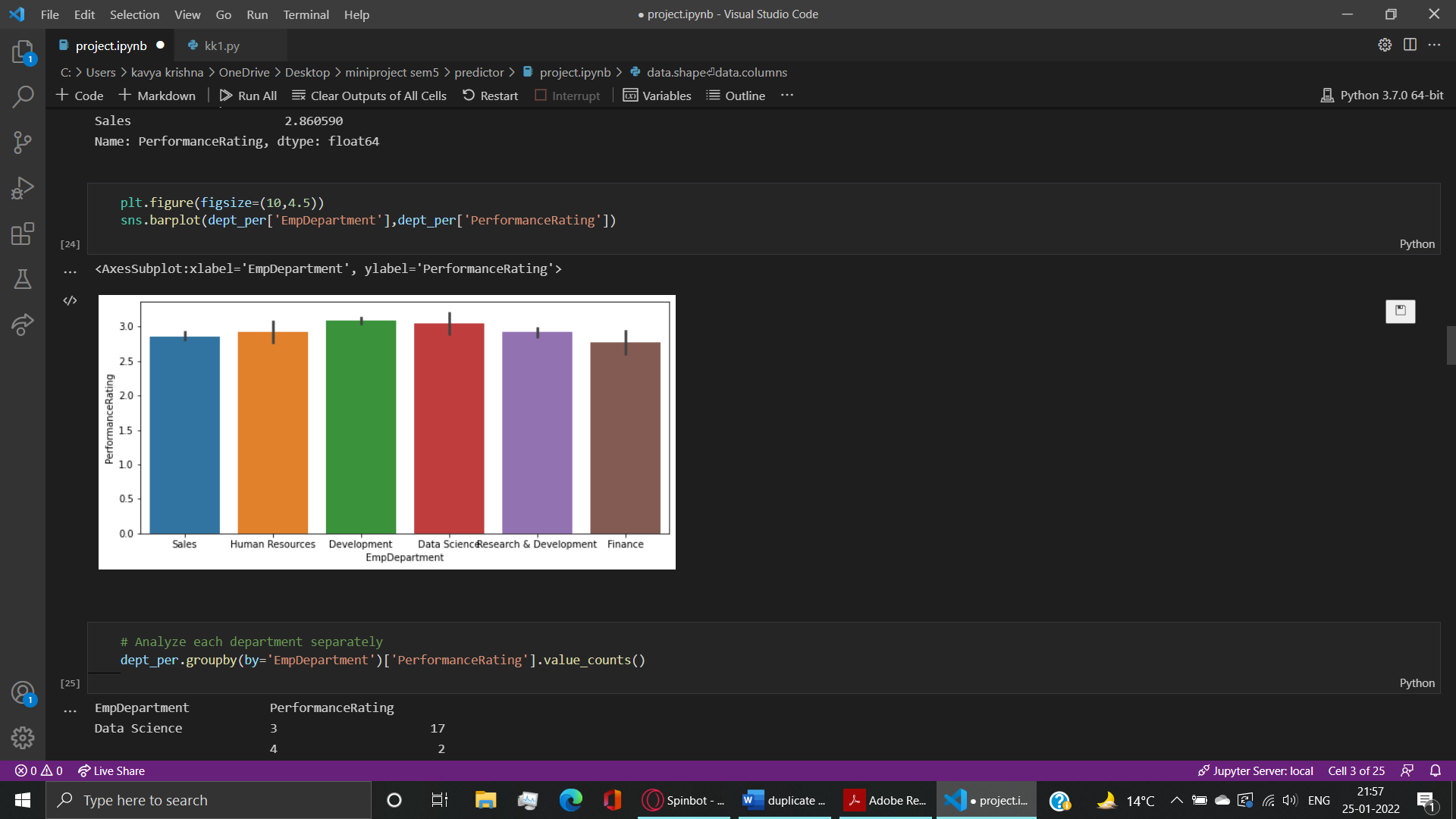


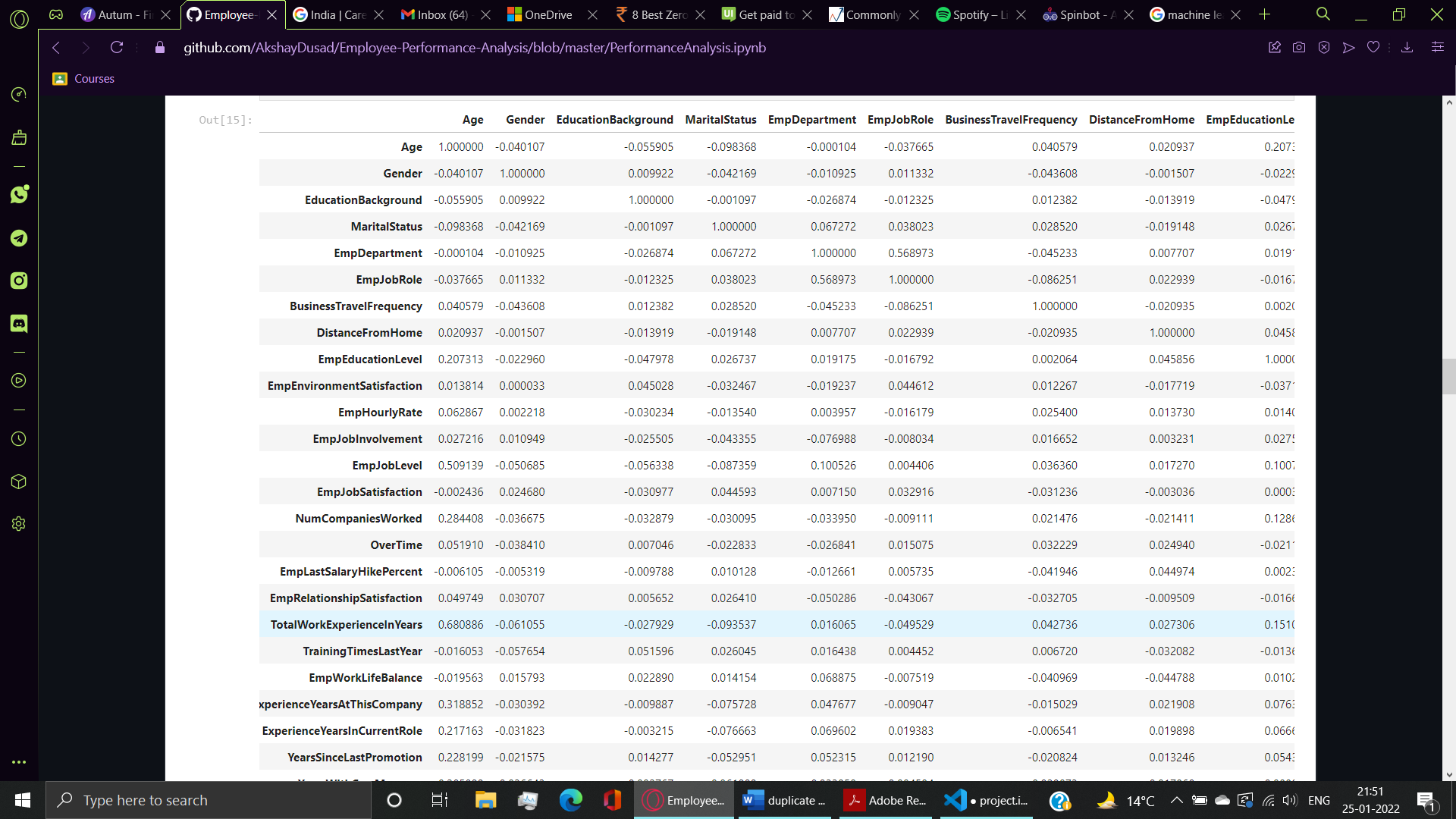


Plotting a separate bar graph for performance of each department using seaborn



**Department wise performance analysis**





**Feature selection:**

There are a ton of segments in the indicator variable. In this way, the connection coeffecient is determined to see which of them are significant and these are then utilized for preparing techniques. From that point, we additionally get the top variables which influence execution. We can see that the main elements selected were Department , Last Salary Hike Percent, Work Life Balance, Job Role, Environment Satisafaction , Experience Years At This Company, Experience Years In Currrent Role, Years Since Last Promotion, Years With Current Manager. These were chosen on the grounds that their connection co-effecient with Performance Rating was more than 0.1.

Normalization and Label Encoding was additionally utilized for include change.

A different consideration thinking about every one of the indicators was done however it brought about diminishing the exactness. Essentially, Principal Component Analysis likewise diminishes the exactness.

Top 3 variables which influence the representative presentation are 1. Employee Environment Satisfaction, 2. Employee Last Salary Hike Percent and 3. Years Since Last Promotion.

After selecting the attributes which actually helps us predict performance of an employee after applying feature selection we have :

EmpDepartment,

EmpJobRole,

EmpEnvironmentSatisfaction,

EmpLastSalaryHike,

EmpWorkLifeBalance,

ExperienceYearsAtThisCompany,

ExperienceYearsInCurrentRole,

YearsSinceLastPromotion,

YearsWithCurrManager,

I trained my model on 70% data and saved 30% for testing. The model is trained on random forest algorithm. I tried training on other algorithms as well but with the amount of data (size of data) and the type of data, random forest performed best with highest accuracy among all with an accuracy of 90.5%.

**CHAPTER 4**

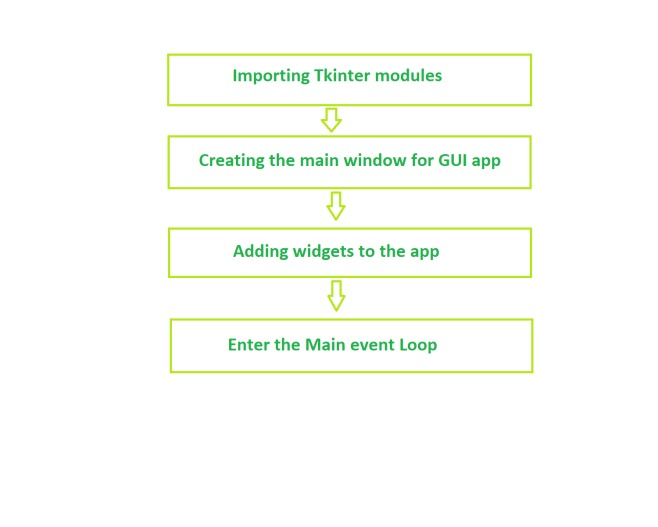
**FUNDAMENTALS OF UI using tkinter**

**4.1 DESIGN GOALS**

This mini project has ensured that the user has an interactive and explorable environment. The interface is user friendly, simple to understand and has tried to ensure that there are no bugs.

Tkinter is a module that is constructed in python which is used to create GUI applications. It is one of the most widely/popularly used modules for developing GUI applications in python as it is simple and easy to work with.

Programmers can create any GUI applications that they desire with the use of Tkinter.

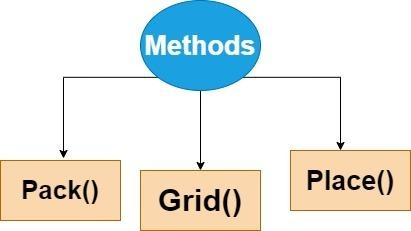


**Figure 4.1: flow of creating GUI**

Widget is a component of Graphical User Interface (GUI) that presentations/delineates data or gives a way for the client to connect with the OS. In Tkinter , Widgets are oobjects ; occasions of classes that address buttons, outlines, etc.

Layout managers additionally are alluded to as geometry managers. They are utilized forsituating, organizing and enlisting gadgets on tkinter window. Python offers 3 design/calculation supervisors.

Tk empowers 3 sorts of geometry managers : placer, packer, & grid.



**Figure 4.2: methods in Tkinter**

**Packer**:

Pack is the ideal to apply of the 3geometrymanagers of Tk and Tkinter. Rather than getting to guarantee precisely in which a gadget need to appear at the show screen, we will guarantee the places of gadgets with the % order comparative with each other. The % order takes care of the subtleties. However the % order is less hard to apply, this organization administrators is obliged in its chances in contrast with the lattice and district troughs.

**Placer**:

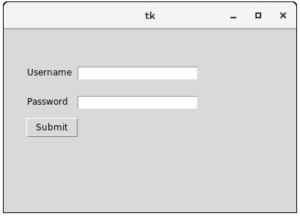
The Place unadulterated arithmetic boss permits you explicitly set the arrangement and length of a window, each in outright terms, or comparative with each unique window. The placement supervisor is likewise gotten to by means of the situation system. it will be distributed to any or every single boundless gadget.

**Grid**:

Grid is in masses of times the awe inspiring decision for best in class use. While rate is sometimes now presently not adequate for changing measurements with inside the design, region gives you entire control of situating every component, except this makes it masses more noteworthy complex than rate and Grid.

**LABELS**

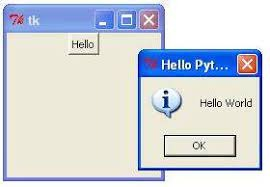
Tkinter Label is a gadget that is utilized to carry out show boxes where you can put text or pictures. The text showed by this gadget can be changed by the designer at any time you need. It is likewise used to perform assignments, for example, to underline the piece of the text and length the text across various lines. It is essential to take note of that a name can utilize just a single textual style at a time to display text.



**Figure 4.3: lables**

**BUTTON**

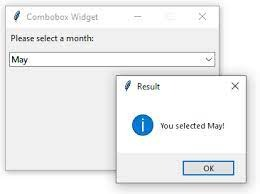
The Button gadget is a customary Tkinter gadget, which is utilized for different sorts of buttons. A button is a gadget which is intended for the client to collaborate with, for example assuming the button is squeezed by mouse click some activity may be begun. They can likewise contain text and pictures like marks. While marks can show text in different textual styles, a button can show text in a solitary textual style. The text of a button can traverse more than one line.



**Figure 4.4: buttons**

**COMBOBOX**

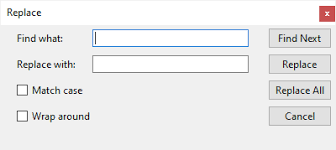
Combobox is a blend of Listbox and a entry field. It is one of the Tkinter gadgets where it contains a down bolt to choose from a rundown of choices. It helps the clients to select as indicated by the rundown of choices showed. At the point when the client taps on the drop-down bolt on the passage field, a spring up of the looked over Listbox is shown down the section field. The chose choice will be shown in the passage field just when a choice from the Listbox is chosen.



**Figure 4.5: combobox**

**FRAME**

An frame is a rectangular locale on the screen. An edge can likewise be utilized as an establishment class to execute complex gadgets. It is utilized to coordinate a gathering of gadgets.



**Figure 4.6: frame**

Each different gadget is a Python object. While making a gadget, you should pass its parent as a boundary to the gadget creation work. The main exemption is the "root" window, which is the high level window that will contain all the other things and it doesn't have a parent.

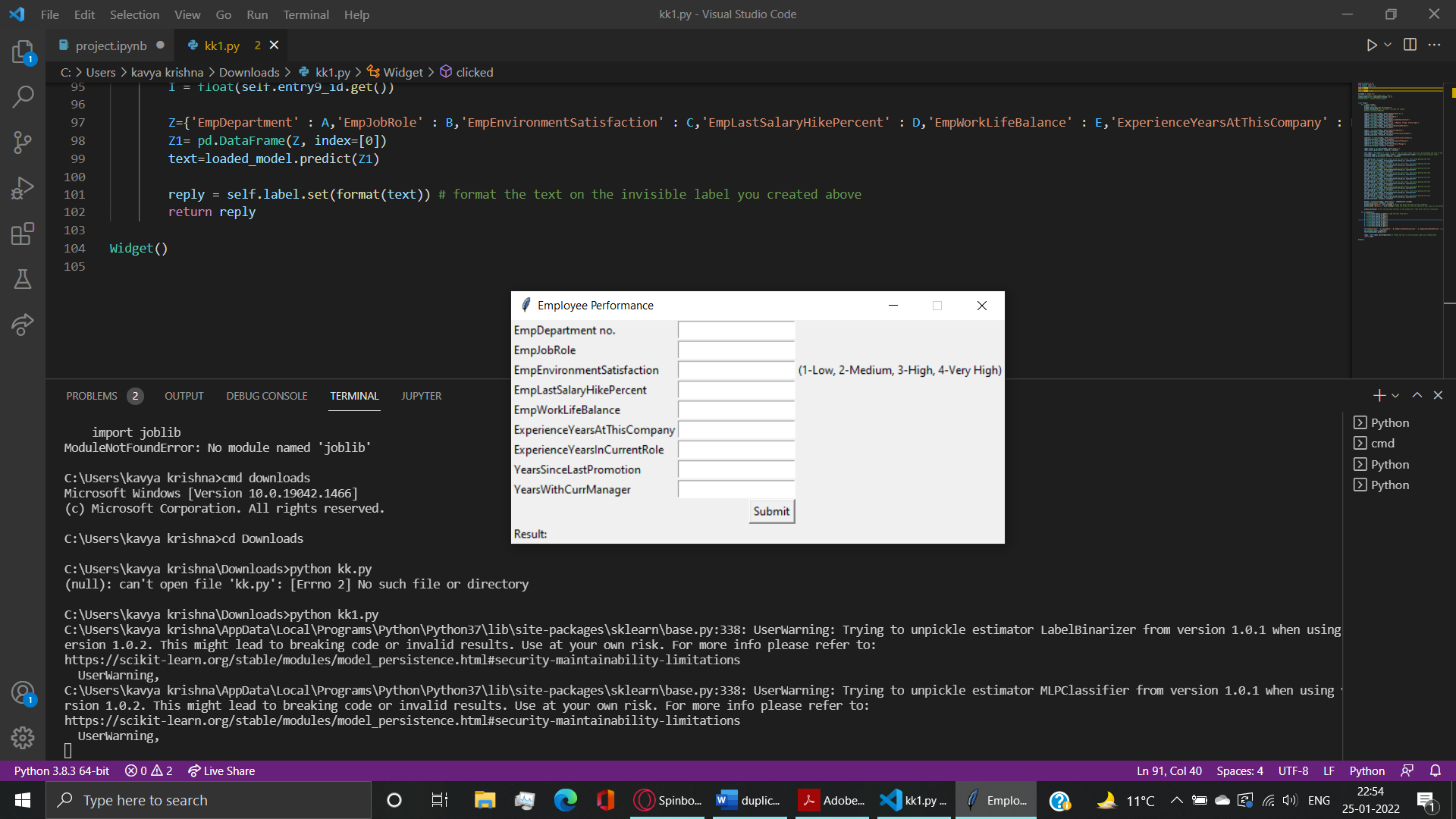
|  |  |
| --- | --- |
| WIDGETS | DESCRIPTION |
| Label | This widget is used to display text or image on the window/frame |
| Button | This widget is used to add buttons to the user interface |
| Canvas | This widget allows one to draw pictures and different types of layouts like texts, graphics etc. |
| Entry | This widget is used to take as input, a single line text entry from user |
| Frame | This widget is used as box or container. It holds and organizes the widgets in an orderly fashion |
| SpinBox | This widget allows users to select from a given number of values |
| ComboBox | This widget contains a down arrow to select from a list of options |
| CheckButton | This widget displays a number toggle buttons which represent various options from which user can select any number of options. |
| RadioButton | This widget is similar to the CheckButton but allows only one option to be selected |
| Scale | This widget is used to provide a slider which allows the user to select any value from the scale |

**Table 4.7: Various widgets available in Tkinter**

**CHAPTER 8**

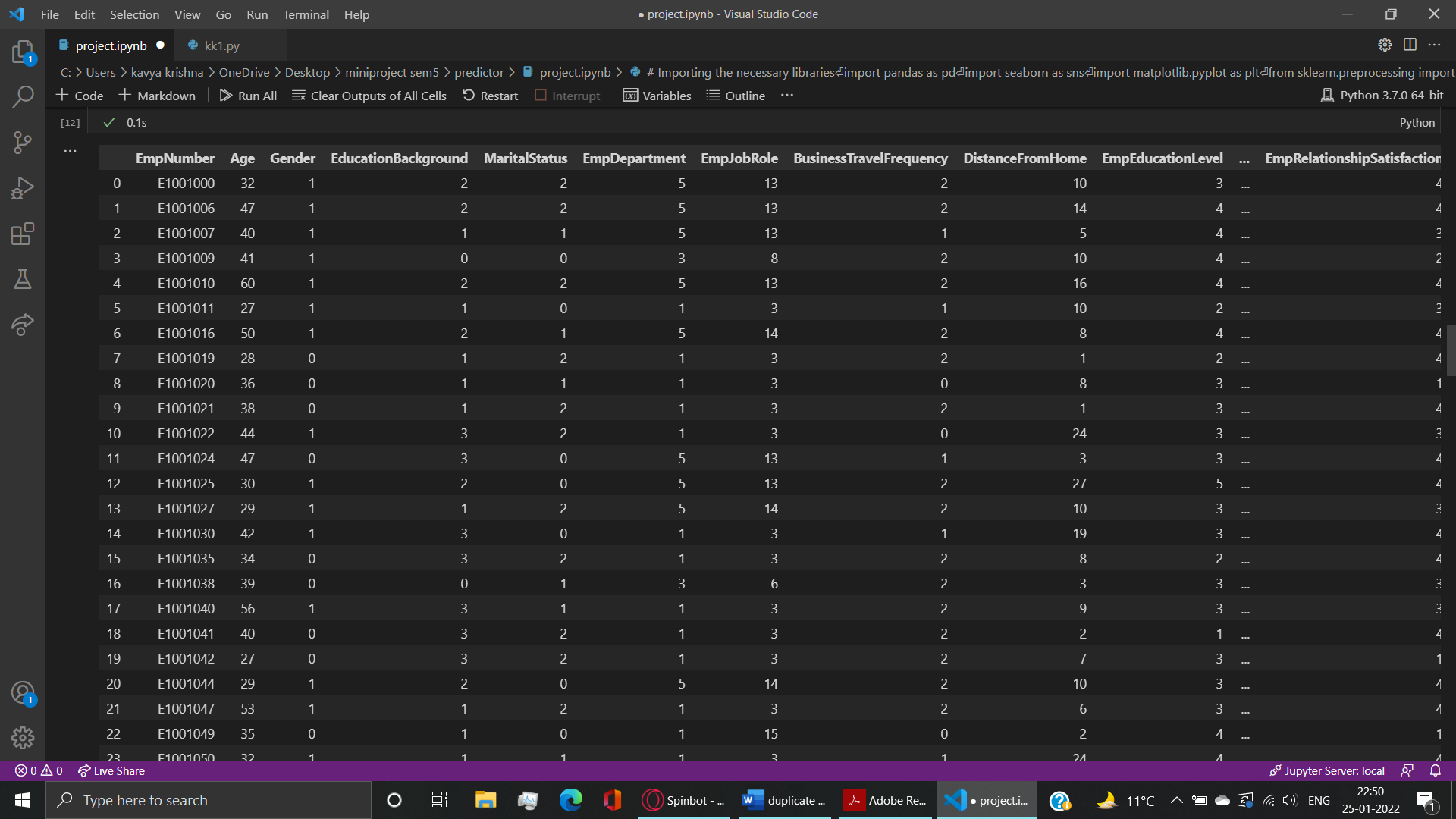
**RESULTS**

**8.1 Launch Screen**



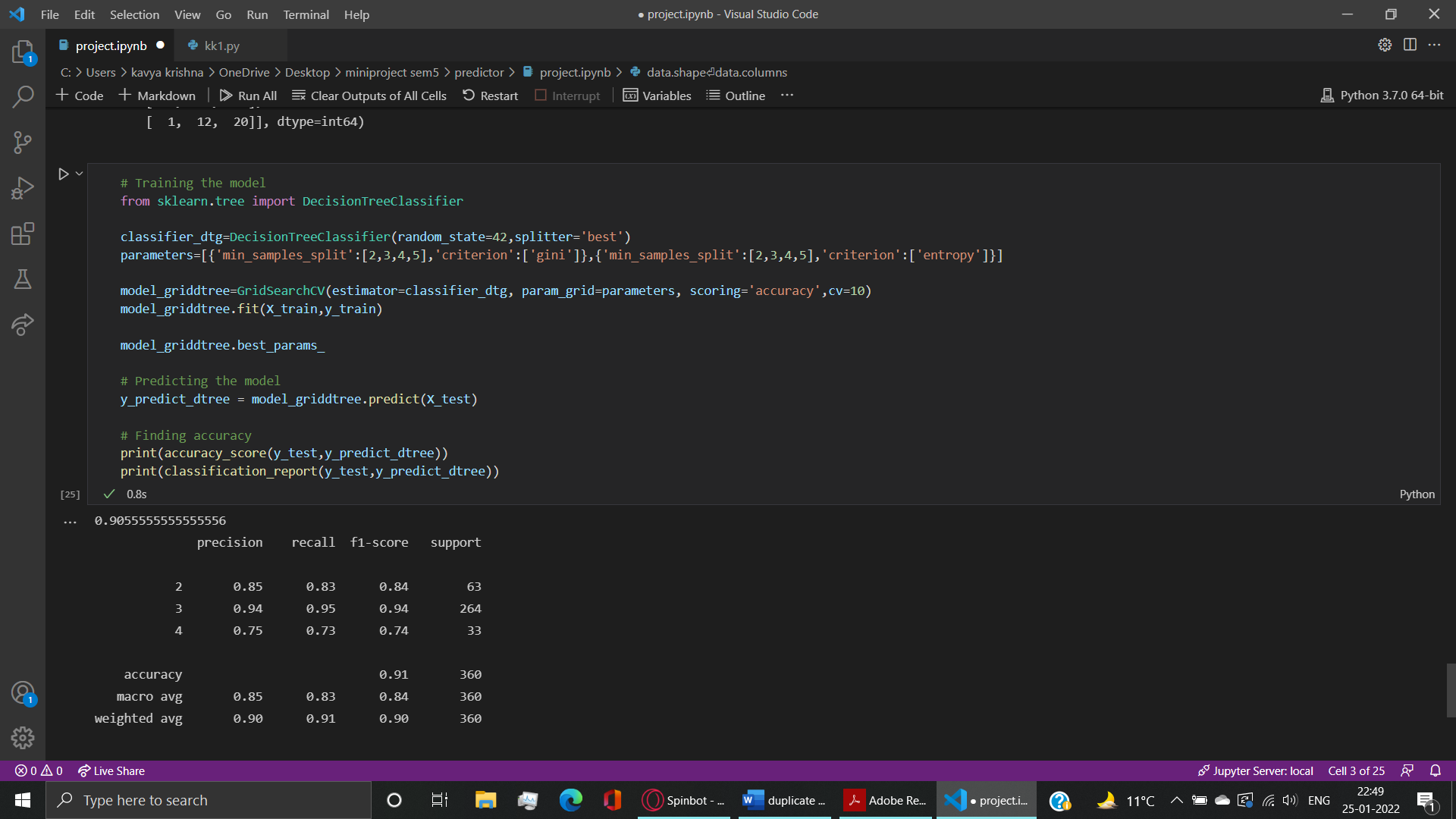
**Figure 8.1: Screen Shot of Launch Screen**

**8.2 Screenshot of data set:**



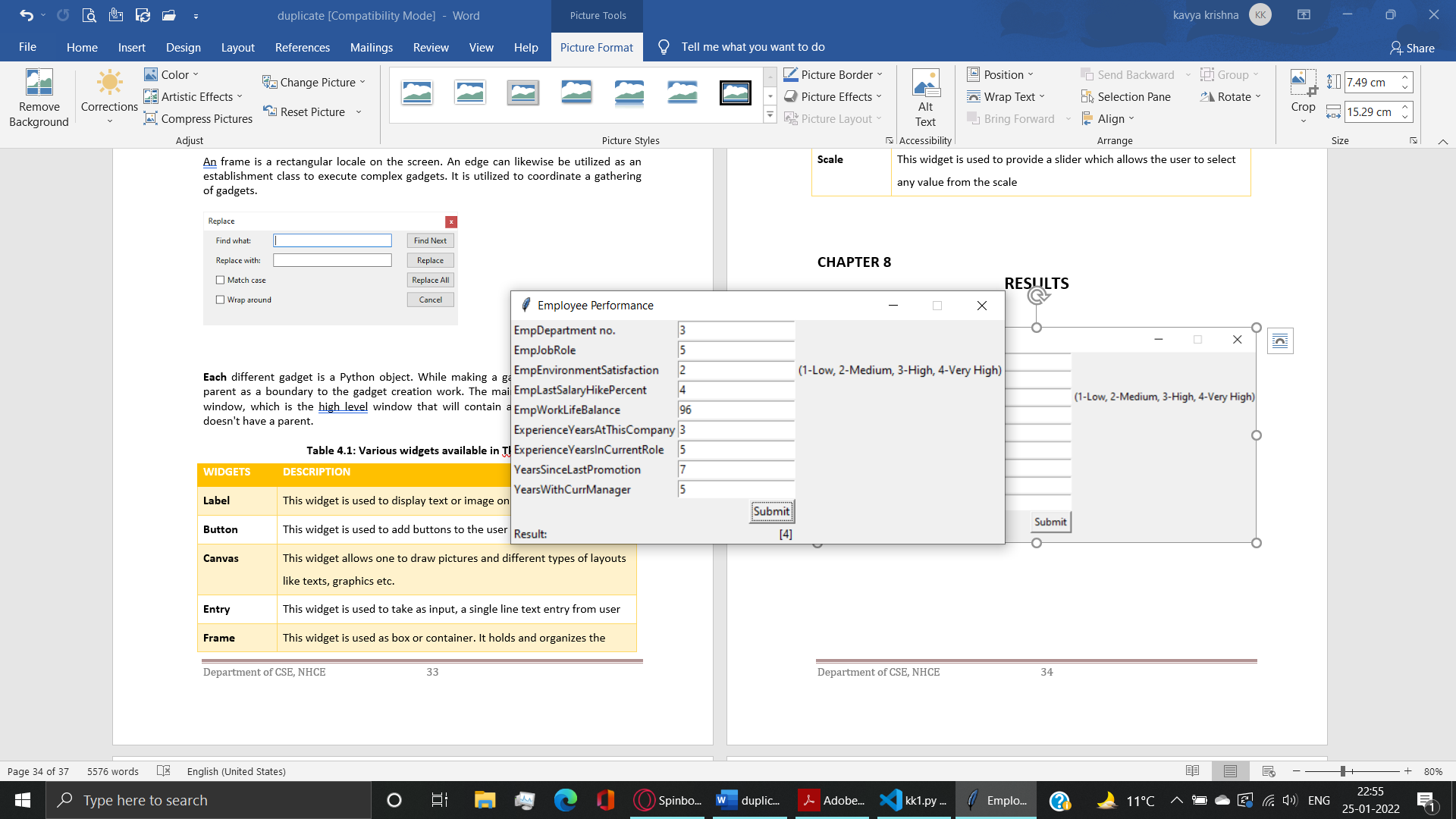
**Figure 8.2: data set**

**8.3 Screenshot of confusion matrix and accuracy:**



**Figure 8.3: Output**

**8.4 OUTPUT**



**Figure 8.4: Output**

**CHAPTER 9**

**CONCLUSION**

The project is successfully completed fulfilling all the requirements of the problem statements. The model works accurately for the provided data and has been trained using random forest algorithm while trying other algorithms as well which really helped me understand machine learning even better. This project has been designed keeping in mind the pace at which whole industry is being automated and this one section, i.e, HR system, which can be automated as well instead of continuing with the manual approach. This helps us save time and utilize it towards more tech evolution.

**REFERENCES**

[1] <https://www.javatpoint.com/> (example for website referred)

{2] <https://docs.python.org/>

[3] A.Conci, J. E. R. de Carvalho, T. W. Rauber, A Complete System for Vehicle Plate Localization, Segmentation and Recognition in Real Life Scene, IEEE LATIN AMERICA TRANSACTIONS, VOL. 7, NO. 5, September 2009. (Example for paper referred)

[4] Joseph Yiu, The Definitive Guide to ARM Cortex-M3 and Cortex M4 Processor, 3rd Edition, Newness Publication (example for book referred)