Methodology:

* Used seaborn to Data Visualization and Exploration
* the perceptron and SVC will be trained on the continuous data (Classification)

Technique :

* Seaborn (To visualization of the data)
* Pandas (For the creating of different data frames and cleaning of data)
* NumPy (Used for mathematical purpose)
* matplotlib (To plot the graphs based on the data)
* Sklearn (To train the models based on the dataset)

Findings :

* Misclassified Samples
* Accuracy score of the model (Perceptron , MLPClassifier , SVC)

Link:

* <https://www.kaggle.com/dan195/classification-of-student-marks>

Methodology :

* Exploratory data Analysis

Technique :

* NumPy ( For linear Algebra )
* Pandas ( For data Processing )
* Matplotlib ( For plotting the graph )

Findings :

* The relation between the raise hand and absence in class. The outcome is that the student who Is absence for more than 7 days he/she has the very low raise hand and the student who is absence for less than 7 days has more.

Link :

* <https://www.kaggle.com/sohailkhan/exploratory-analysis-1>

Methodology:

* A benefit of EDA is that is allows you to examine the data as it is without making any assumptions
* Just Analyzing the data in the dataset to understand the relation between the data to each other and many more.

Technique :

* NumPy ( Used for the Linear Algebra )
* Pandas ( Used for the data processing )

Findings :

* By applying such methods they have plotted the graphs and based on that we have to understand the insight of the data.
* They have compared many fields with each others to identify the correlation between them like Raise hand → Discussion , Announcement → raise hand , Discussion → Visit Resources etc.…

Links :

* <https://www.kaggle.com/wittykitty/exploratory-data-analysis-for-beginners>

Methodology :

* EDA ( Exploratory data Analysis )
* Sklearn to predict the model values on the dataset

Technique :

* NumPy (Used for the Linear Algebra )
* Pandas ( Used for data preprocessing )
* ggplot( To plot the data )
* Sklearn ( For model execution and finding the accuracy )

Findings :

* By applying the EDA he has find all the relationship of the data each and every possible combination of data with each other.
* By applying this in the histogram , tilegraph , boxplot (With every possible combination) and many other kind of the graph to visualize the data in the appropriate way.
* He has even visualized the data in the Decision tree and train the models like Decision tree , Random Forest , Support Vector Machine and at the end he has used Ensemble Learning to combine all the three models output and learn from that.
* At the end he has find the accuracy for that. Which is 88.43 %

Link :

* <https://www.kaggle.com/arathee2/predicting-classes/report>

Methodology :

* Use Machine Learning to predict the student Performance on the various data like raise hand etc..

Technique :

* NumPy (Used for the Linear Algebra )
* Pandas ( Used for data preprocessing )
* Matplotlib ( Used to plot the graphs )
* Sklearn ( For Machine Learning models executions )

Findings :

* Plot the graph for the pivot table
* Plot the graphs for correlation of data with each other
* Plot other Histograms for better data visualization of data based , To identify the which data column is important and which is not
* Used Machine Learning Algorithms like Random Forest , Logistic Regression , SVM , stratified KFold for KFold validation to get the better accuracy and many more algorithms.
* Then plot the graph for all the algorithms accuracy score to identify easily.
* Again after some modification in the dataset using transformation he has applied All the algorithms again and find the accuracy score for that all.
* And again plot the graph for that algorithm again and get the better result.

Link:

* <https://www.kaggle.com/rmalshe/student-performance-prediction>

Methodology :

* Find the prediction rate of the Logistic Regression and plot that accuracy in the Confusion metrics.

Technique :

* NumPy ( Used for the Linear Algebra )
* Pandas ( Used for the Processing of data )
* Sklearn ( To implement the Logistic Regression & Plot the confusion metrics )

Findings :

* Accuracy score of the Linear Regression
* Plot the confusion metrics

Links:

* <https://www.kaggle.com/jpotts18/logistic-regression-1>

Methodology :

* Correlation of data using Machine Learning Algorithm
* PCA ( Principle Component Analysis )

Technique :

* Numpy ( For Linear Algebra )
* Pandas ( For Processing of data )
* Seaborn ( To visualize the EDA )
* Matplotlib ( To plot the graph of different type )

Findings :

* The correlation graph between different student attributes
* PCA graph to visualize the data ( Individual explained Variance )

Links :

* <https://www.kaggle.com/isaacinyang/notebookf9a5a17c71>

Methodology :

* The bar chart to display the nationality of student from the dataset

Techniques :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Matplotlib ( For data visualization into the graph )

Findings :

* The graph that represent how many students are from which country in the histogram graph and Pie chart.

Link :

* <https://www.kaggle.com/shashanksingh/charting-nationality>

Methodology :

* Visualize the data in the histogram to with each other to understand the data
* Plot the Heatmap of correlation with the help of seaborn.

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Matplotlib ( For visualizing the data in graph )
* Seaborn ( For visualizing the heap map of correlation )

Finding :

* A histogram of Importance of feature to identify the feature which is impacting more on the training dataset.
* A heat map of correlation which is identify the correlation of the features with each other. The highly correlated the high impact on the training dataset.

Link:

* <https://www.kaggle.com/jeffreynghm/academic-performances>

Methodology :

* Data Exploration

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )

Findings :

* All the classes available in the dataset
* All the classes with the number of occurrence in the dataset

Link :

* <https://www.kaggle.com/jjconde/quick-data-exploration>

Methodology :

* Random Forest and Ensemble Learning Algorithm to find the Accuracy.
* Used the Hyperparameter to find the appropriate parameter for that dataset according to the model.
* Feature selection to find the important feature from the dataset.

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Sklearn ( To implement the Algorithm )
* Used L1-based Feature selection

Findings :

* The accuracy score of the Random Forest classifier and Bagging classifier with hyperparameter tuning and without hyperparameter tuning

Link:

* <https://www.kaggle.com/chandiniunnikrishnan/student-grades-classification>

Methodology :

* Data Exploration
* Machine Learning Algorithm
* Cross-Validation for hyper parameter tuning

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data Processing )
* Matplotlib ( For Visualizing the data in the graph )
* Sklearn ( For Machine Learning )

Findings :

* Best Hyper-parameter for Random Forest Classifier
* Feature important table represent the feature importance in the dataset.
* The graph of importance of the feature in a form of histogram

Link:

* <https://www.kaggle.com/gemunu/student-perfomance/notebook>

Methodology :

* Data Analysis
* Data Visualization

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Matplotlib ( For data visualization in graphs )

Findings :

* The heat map of correlation between raise hand , Visit Resource , Announcement , Discussion.
* Seaborn to represent the graphs in different ways.
* The graph of Discussion -> raise hand , Visit Resource , Announcement , Discussion

Link:

* <https://www.kaggle.com/liaopan/data-analysis-in-students-academy>

Methodology :

* Regression used to predict the scores of the students
* GridSearchCV to find the hyperparameter of the respective model.
* kFold validation to find the Folds the dataset

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Sklearn ( For Regression and Hyperparameter tuning )
* Matplotlib ( For visualizing the data )

Findings :

* The best hyper-parameter for the Regression algorithms
* And the prediction score of all the model based on the dataset

Link:

* <https://www.kaggle.com/raghavrastogi75/student-scores-analysis-and-prediction/script>

Methodology :

* Decision Tree and Random Forest
* EDA

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Seaborn ( For EDA )
* Matplotlib ( For Visualizing graphs )
* Sklearn ( For Decision Tree and Random Forest )

Findings :

* Confusion matrix for the Decision Tree and Random Forest
* Heat map of the Classification Report
* Accuracy score for the Classification

Link:

* <https://www.kaggle.com/miya0930/decisiontree-and-randomforest>

Methodology :

* Data Analysis
* Machine Learning ( Random Forest Classifier , SVC and Decision Tree )

Technology :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Matplotlib ( For data visualization )
* Sklearn ( For Random Forest )

Findings :

* Graph of raise hand -> Class , class -> Parent , class -> semester , class -> Student Absence , class -> Gender , class -> Parent school satisfaction and many more…
* Accuracy Score for the Decision Tree , Random Forest and SVC.

Link :

* <https://www.kaggle.com/ritgoe/student-data-analytics>

Methodology :

* EDA ( Exploratory Data Analysis )
* Logistic and XGBoost

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Matplotlib ( For data visualization )
* Sklearn ( For Linear and XgBoost )
* Seaborn ( For EDA )

Findings:

* Get the graph of gender , Place of Birth , student Absence Days etc…
* Logistic Regression results , score and classification Report and same for XGBoost.
* The graph of important feature

Link :

* https://www.kaggle.com/rahulvks/exploratory-data-analysis-logistic-and-xgboost

Methodology :

* Data visualization
* Seaborn

Technique :

* Pandas ( For data processing )
* Seaborn ( For data visualization )

Findings :

* Count Plots of every feature available in the dataset with horizontal and vertical graph

Link :

* <https://www.kaggle.com/slamnz/seaborn-tutorial-count-plots>

Methodology :

* EDA
* Machine Learning ( Decision Tree and Random Forest Classifier )

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Sklearn ( For Machine Learning )
* Seaborn ( For EDA )

Findings:

* The graphs which represents the data in the very good manner
* The Accuracy score for the Decision Tree Classifier and Random Forest Classifier

Link :

* <https://www.kaggle.com/ritgoe/notebookc795091ab7>

Methodology :

* Data visualization
* Tree Bases classification

Technique :

* NumPy ( For Linear Algebra )
* Panada ( For data processing )
* Matplotlib ( For data visualization )
* Sklearn ( For Logistic Regression and Decision Tree )

Findings :

* The graphs for the different types of comparison of the data with each other
* Done the data exploration
* Get the accuracy score , Classification Report , Confusion metric for the Random Forest and Logistic Regression
* Done the Gradient Boosting for the getting the better result.
* By adding some more hyper-parameter find the better accuracy model for the Gradient Boosting

Link:

* <https://www.kaggle.com/rruizendaal/visualizations-tree-based-classification-methods>

Methodology :

* Machine Learning ( Find best classification algorithm for the dataset )

Technique :

* Pandas ( For data processing )
* Sklearn ( For Classification Algorithms )

Findings :

* All the classification algorithms Average score and Standard Deviation to find which is best of all them

Link :

* <https://www.kaggle.com/slamnz/what-is-the-best-classifier-to-start-with>

Methodology :

* Performance analysis of the student
* Cor-relation of the data with each other
* Machine Learning ( perceptron , SVC , MLP Classifier , Random Forest and XGBoost )

Technique :

* NumPy ( For Linear Algebra )
* Pandas ( For data processing )
* Sklearn ( For Machine Learning )
* Seaborn ( For EDA )

Findings :

* Histogram graph for the different data available in the dataset
* Heat Map of co-relation of data available in the dataset
* Regression plot of the dataset
* Classification Report , Accuracy score for the Machine Learning Algorithm
* The Tile graph for the feature importance according to XGB ( XGBoost ) algorithm.

Link:

* <https://www.kaggle.com/satadru5/student-peroformance-analysis>