**Baseball Case Study**

**INTRODUCTION:**

The project comprises data from2014 Major League Baseball seasons in order to develop an algorithm that predicts the number of wins for a given team for 2015 season based on certain indicators of success. Total 16 different features are given and used as inputs to the machine learning and based on these output will be predicted which are nothing but value that represents the numbers of wins.

**16 Inputs given in the project (Baseball Case Study)**

1.W- This indicates the number of wins credited to a pitcher.

2.R-This indicates Runs scored.

3.AB-This means At bat or time at bat.

4.H-This means Hit. Its also called a “base hit”.

5.2B-This means act of a batter striking the pitched ball and safely reaching second base without being called out by the umpire.

6.3B- This means a Tripple. It’s the act of batter safely reaching third base after hitting the ball. Its sometime called a “ three-bagger” or

“ three-base hit”

7.HR-This means Home runs. Its scored when the ball is hit in such a way that the batter is able to circle the bases and reach home plate safely in one play without any error.

8.BB-This means Base on balls( also called a “ walk”)

SO-Also denoted as “K” means strikeout.

9.SB-This means stolen house.

10.RA-This means Run Average.

11.ER-This means Earned Run.

12.ERA-This means Earned Run Average.

13.CG-This means completeb Game.

14. SHO-This means Shutout.

15.SV - This means Save. It's credited to a pitcher who finishes a game for the winning team under certain prescribed circumstances: number of games where the pitcher enters a game led by the pitcher's team, finishes the game without surrendering the lead, is not the winning pitcher, and either (a) the lead was three runs or fewer when the pitcher entered the game; (b) the potential tying run was on base, at bat, or on deck; or (c) the pitcher pitched three or more innings.

16.E - This means Errors. It's an act, in the judgment of the official scorer, of a fielder misplaying a ball in a manner that allows a batter or baserunner to advance one or more bases or allows a plate appearance to continue after the batter should have been put out. The term error is sometimes used to refer to the play during which an error was committed: number of times a fielder fails to make a play he should have made with common effort, and the offense benefits as a result

**PREDICTION** :Number of wins for given team for 2015 based on the above indicaters.

**HOW WILL IT BE HELPFUL FOR THE BUSINESS:** we can use these kind of model for different teams and sports which can predict their future perfomances.

**HOW DID I START THIS PROJECT:** First we got the data from GitHub and we downloaded the data and we loaded our data to Jupiter notebook. Following that we performed cleaning of our data after that we performed preprocessing and we analysis data in detail , we saw density plot the we performed features selection. Moreover we observed each and every features and we dropped the features which were not good and we selected features which were giving good results. Additionally we trained our model and begins with Linear Regression , Random forest and Xgboost Regressor ,hence we observed that Random forest was performing excellent among all models.

**GRAPH**: In Graph we used Plot and density distribution for our target variable which is win , that shows our data is distributed normally following that we plotted a heat map between target variable and other dependent variable. Hench this depicts correlation among our features.

**MODELS USED IN PROJECT:**

1. **Linear Regression**
2. **Random Forest**
3. **Xgboost**

**FINAL MODEL SELECTED FOR PREDICTION:** we selected Random forest as it scored highest among all models.

**CONCLUSION WITH BENEFIT**: It will help the team for their better performance and decision making in future.

**AVOCADO PROJECT**

**INTRODUCTION:**

This data was downloaded from the Hass Avocado Board website in May of 2018 &amp; compiled

into a single CSV.

The table below represents weekly 2018 retail scan data for National retail volume (units) and

price. Retail scan data comes directly from retailers’ cash registers based on actual retail sales

of Hass avocados.

Starting in 2013, the table below reflects an expanded, multi-outlet retail data set. Multi-outlet

reporting includes an aggregation of the following channels: grocery, mass, club, drug, dollar

and military. The Average Price (of avocados) in the table reflects a per unit (per avocado) cost,

even when multiple units (avocados) are sold in bags.

The Product Lookup codes (PLU’s) in the table are only for Hass avocados. Other varieties of

avocados (e.g. greenskins) are not included in this table.

Some relevant columns in the dataset:

 Date - The date of the observation

 AveragePrice - the average price of a single avocado

 type - conventional or organic

 year - the year

 Region - the city or region of the observation

 Total Volume - Total number of avocados sold

 4046 - Total number of avocados with PLU 4046 sold

 4225 - Total number of avocados with PLU 4225 sold

 4770 - Total number of avocados with PLU 4770 sold

**Inspiration /Label**

The dataset can be seen in two angles to find the region and find the average price .

Task: One of Classification and other of Regression

Do both tasks in the same .ipynb file and submit at single file.

**PREDICTION** : Find the Region and find and the Average price.

**HOW WILL IT BE HELPFUL FOR THE BUSINESS:** It will help the company to get the best average price and the region where we can get the best Avocado. It will help firm to take the best decision in the future as well.

**HOW DID I START THIS PROJECT**: First we got the data from GitHub and we downloaded the data and we loaded our data to Jupiter notebook. Following that we performed cleaning of our data after that we performed preprocessing and we analysis data in detail ,after that we did some EDA , in that we plotted density distribution for our target variable i.e Average price, here we had to conclude two task .

1. **Regression problem:**

**For this target variable was Average price**

1. **Classification problem:**

**For this target variable was Region**

Regression: Here we plotted normal distribution , we observed average price was closed to normal distribution followed that we plotted some other graphs:

1. **Type of Avocado( i- conventional, ii- organic)**

Both were equal in numbers after that we performed some preprocessing in which we removed outliers from our dataset using IQR method. Moreover we plotted heatmap on our dataset to find out correlation between different features of our dataset with the help of we dropped some least important or zero correlated features and chosen best features instead. Addionally we performed some regression task in which our target variable was Average price , here we started with Random forest which gave good result later on we performed hyper parameter tuning , hench we got best model. For this model R square = 0.77 and MSE= 0.03.

Classification: After regression we performed classification and here we did follow same steps/task and for this target variable was ‘Region’ , we trained the different model again

1. **Random Forest Classic fire**
2. **Decision Tree Classic fire**

Based on above observation we concluded that Random forest was giving best output hence got chosen.

**MODELS USED IN PROJECT**:

1.Random Forest

2.Decision Tree Classical fire

**FINAL MODEL SELECTED FOR PREDICTION:** Random fore

**HR Analytics Project:**

**Understanding the Attrition in HR**

**INTRODUCTION**:The given project is about a lot companies which hire number of employees. Companies invest a lot of money and time and conduct programme to enhance the manpower of the organisation. Attrition in human resources refers to the gradual loss of employees overtime. In general, relatively high attrition is problematic for companies. HR professionals often assume a leadership role in designing company compensation programs, work culture, and motivation systems that help the organization retain top employees.

How does Attrition affect companies? and how does HR Analytics help in analyzing attrition? We will discuss the first question here and for the second question, we will write the code and try to understand the process step by step.

**WHAT IS ATTRITION**:  Employee attrition is defined as employees leaving their organizations for unpredictable or uncontrollable reasons. Many terms make up attrition, the most common being termination, resignation, planned or voluntary retirement, structural changes, long-term illness, layoffs.

**REASONS FOR ATTRITION:** Attrition, or employee turnover, can be influenced by a variety of factors. Here are some common reasons:

1. Lack of Career Advancement: Employees may leave if they feel there's limited opportunity for growth or promotion within the company.
2. Inadequate Compensation: If employees believe they can earn more elsewhere or feel their compensation does not reflect their skills and contributions, they may seek opportunities elsewhere.
3. Poor Work-Life Balance: Excessive workloads, long hours, or inflexible schedules can lead employees to leave in search of a better balance.
4. Unhealthy Work Environment: A toxic work culture, poor management practices, or lack of support can drive employees away.
5. Lack of Recognition: Employees who feel their efforts and achievements are not acknowledged or valued might seek recognition elsewhere.
6. Job Misalignment: If the job role or company culture doesn’t align with an employee’s expectations or values, they might look for a better fit.
7. Limited Training and Development: Employees may leave if they feel they’re not receiving the training or development needed to stay current in their field or advance their careers.
8. Better Opportunities: Sometimes employees leave for new challenges, higher pay, or more attractive benefits offered by other organizations.
9. Personal Reasons: Factors such as relocation, changes in personal circumstances, or health issues can also contribute to attrition.
10. Leadership Issues: Ineffective or unsupportive leadership can drive employees to leave, especially if they feel undervalued or unsupported by their managers.

Addressing these issues proactively can help reduce turnover and improve overall employee satisfaction.

**PREDICTION** : we will predict two question here.

1. How does Attrition affect companies?
2. How does HR Analytics help in analyzing attrition?

**HOW WILL IT BE HELPFUL FOR THE BUSINESS**:It will help companies to retain its employees and gain the profit in long run. It also helps HR department in giving appraisal to its employee.

**HOW DID I START THIS PROJECT:** Same as above project we got the data from GitHub and we downloaded the data and we loaded our data to Jupiter notebook. Following that we performed cleaning of our data after that we performed preprocessing and we analysis data in detail ,after that we performed some EDA then we found out in the given data there are total 882 male and 588 female, their education level was between 1-5 and most of the employees were at lever 3 along with this information their marital status was given out of which 470 were married and 377 were divorced ,we did some basic kind of plotting and later on we did some feature selection . Total 34 features were given in the data out of which we dropped some after detailed observation and we selected few based on their performances. Moreover we trained some model and we used decision tree, random forest, classifire and xgboost classifire. Later on we used four parameter i.e accuracy, precision, recall and F1. All the model were giving good result. We selected random forest because it was giving best result, again we did some basic hyper parameter tuning with the help of that we got our final model, here all our metrices were 80% which is quite good. Hence we also plotted ROC curve , which shows our model is performing is better.

**MODEL SELECTED FOR PREDICTION:**

1. **Decision Tree**
2. **Random Forest**
3. **3. Classic fire**
4. **Xgboost classifire**

**FINAL MODEL SELECTED FOR PREDICTION:** Random Forest

**CONCLUSION WITH BENEFITS**: It helps HR to take right decision

While attrition can present challenges, it also offers some potential benefits for HR and organizations. Here are a few positive aspects:

1. **Opportunity for Fresh Talent:** Attrition allows organizations to bring in new talent with fresh perspectives and up-to-date skills, which can drive innovation and growth.
2. **Cost Savings on Stagnant Positions:** Sometimes, attrition helps eliminate roles that may have become redundant or less critical, allowing the organization to reallocate resources more effectively.
3. **Chance to Reassess Roles and Responsibilities:** Departures provide an opportunity to review and potentially restructure roles to better align with current business needs and strategies.
4. **Potential to Enhance Employee Engagement:** When managed well, turnover can lead to improvements in workplace culture and employee engagement, especially if it triggers positive changes in management practices or work conditions.
5. **Opportunity to Strengthen Hiring Processes:** High turnover can prompt HR to refine recruitment processes, ensuring better alignment between candidates' skills and organizational needs.

**THANK YOU**