PROJECT REPORT

(Project Term August-November 2021)

Prediction of Hotel Booking-Cancellation

Submitted by

Pratyush Pranjal **Registration Number: 11906033**
Nishant **Registration Number: 11910329**

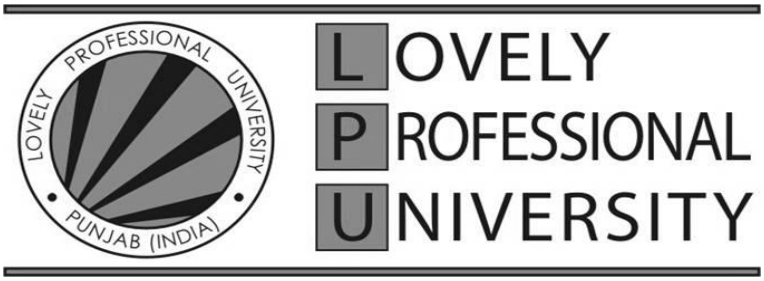
Project Group Number …….

Course Code INT246

Under the Guidance of

Dr. Sagar Pande

School of Computer Science and Engineering



DECLARATION

We hereby declare that the project work entitled **Prediction of Hotel Booking-Cancellation** is an authentic record of our own work carried out as requirements of Project for the award of B.Tech degree in CSE from Lovely Professional University, Phagwara, under the guidance of Dr. Sagar Pande , during August to November 2021. All the information furnished in this project report is based on our own intensive work and is genuine.

Project Group Number: ……

Name of Student 1: **Pratyush Pranjal**
Registration Number: **11906033**

Name of Student 2: **Nishant**
Registration Number: **11910329**

CERTIFICATE

This is to certify that the declaration statement made by this group of students is correct to the best of my knowledge and belief. They have completed this Project under my guidance and supervision. The present work is the result of their original investigation, effort and study. No part of the work has ever been submitted for any other degree at any University. The Project is fit for the submission and partial fulfillment of the conditions for the award of B.Tech degree in CSE from Lovely Professional University, Phagwara.

Dr. Sagar Pande

Designation

School of Computer Science and Engineering,
Lovely Professional University,
Phagwara, Punjab.

Date : 20 November 2021

ACKNOWLEDGEMENT

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to Mr. Sagar Pande for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents, my family and friends and all the google scholar for their kind co-operation and encouragement which help me in completion of this project.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

Project Group Number: ……

Name of Student 1: Pratyush Pranjal

Registration Number: 11906033

Name of Student 2: Nishant

Registration Number: 11910329

Abstract

Nowadays, there are multiple factors that affect a hotel's business.

But in this project, we are only concerning ourselves with booking cancellations that affect its business at a great level. Normally, hotel thrive on online bookings where customers don't even need to physically book the reservation. There is a great abundance of online facilities that aid this process and make sure that anyone can book reservations from any part of the planet.

Throughout this ML model, we work on the sole purpose of looking for patterns where we can derive a particular pattern on the

We have predicted the cancellations majorly using some features:-

lead_time, total_of_special_requests, required_car_parking_spaces, booking_changes and previous_cancellations.

Heart disease is one of the major causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the field of clinical data analysis. With the advanced development in machine learning (ML), artificial intelligence (AI) and data science has been shown to be effective in assisting in decision making and predictions from the large quantity of data produced by the healthcare industry

ML approaches has brought lot of improvements and broadens the perspective that people have for the business world, here, in terms of hotel features that affect bookings. One of such factors is looking for pattern among customers and predicting which new particular customer is most likely to cancel the booking. Initially ML was used to find patterns and on the basis of such patterns, we predict the outcomes for future. There are many features/factors that lead to cancellation of booking which can be worked upon later on.

In this paper we propose a method to finding important features by applying machine learning techniques. The work is to design and develop prediction of booking cancellations by feature ranking machine learning. Hence ML has huge impact in the business world although here it its restricted towards hotel booking cancellation but overall, everywhere, and also drive complex decisions and to create innovative products for businesses to achieve key goals.

TABLE OF CONTENTS

1. Introduction

One of the biggest problems and challenges facing the hospitality industry is the significant number of canceled reservations. Common reasons for cancellations include a sudden deterioration in health, accidents, bad weather conditions, schedule conflicts, or unexpected responsibilities. Interestingly, a noticeable group consists of customers who, after making a reservation, are still looking for new, better offers, and even make many reservations at the same time to be able to choose the most advantageous one.

The use of machine learning to forecast and identify potential cancellations is also playing an increasing role. There are many systems to support hotel management that use booking data. Various machine learning algorithms are used for this purpose, ranging from support vector machines, through artificial neural networks, to the most common tree-based models.

The exploration of trained models should be treated as one of the key factors in the design of hotel management support systems. Business validation and ethical verification of solutions are necessary. Bearing in mind that a strict cancellation policy or overbooking strategy can have negative effects on both reputation and revenue, systems designers should be wary of unfairly biased behavior. At the same time, the use of explanatory artificial intelligence methods is helpful in creating models with better performance scores.

2. PROBLEM STATEMENT:

Over the years, the hotel industry has changed with a majority of bookings now made through third parties such as Booking.com (source). Those Online Travel Agengies (OTA) have transformed cancellation policies from a footnote at the bottom of the page to the main selling point in their marketing campaigns (source). As a result, customers have become accustomed to free cancellation policies. In fact, a study conducted by D-Edge Hospitality Solutions found that cancellation rate across all channels has risen by 6% over the past four years, reaching almost 40% in 2018 (source). This increase in booking cancellation makes it harder for hotels to accurately forecast, leading to non-optimized occupancy and revenue loss (source).

When hotels try to protect themselves by using services such as Booking.com's "Risk Free Reservations", the burden then falls on OTAs. Indeed, this service requires the OTA to pay for the reservation if the booking is canceled and they cannot find a new guest to occupy the room (source). One thing is clear, whether you are a hotel or an OTA, cancellations have an negative financial impact on your business.

In addition to the direct financial consequences of cancellations, they also cause operational problems (such as over or understaffing). Those problems may lead to decrease customer satisfaction and negative reviews. In a world where more and more customers check online reviews before picking a hotel, those reviews can have major impacts. Indeed, TripAdvisor's reviews and scores influenced around $546 billion of travel spending during 2017 (source). At a single hotel level, an increase in online reputation score has been linked to an increase in occupancy and revenue (source). We can clearly understand why avoiding bad reviews due to a room not being ready when the guest arrives can be very valuable for a business. This requires knowing which booking to prioritize. It is therefore very useful for hotels to know which bookings are likely to get canceled in order to plan their operations accordingly.

Characteristics of the booking itself may be good indicators of whether or not a booking will be canceled. For instance, the average length of stay of canceled reservations is 65% higher than non-canceled booking, with a lead time of 60 days (source). Engaging with the reasons why people are cancelling and what types of bookings are being canceled is crucial.

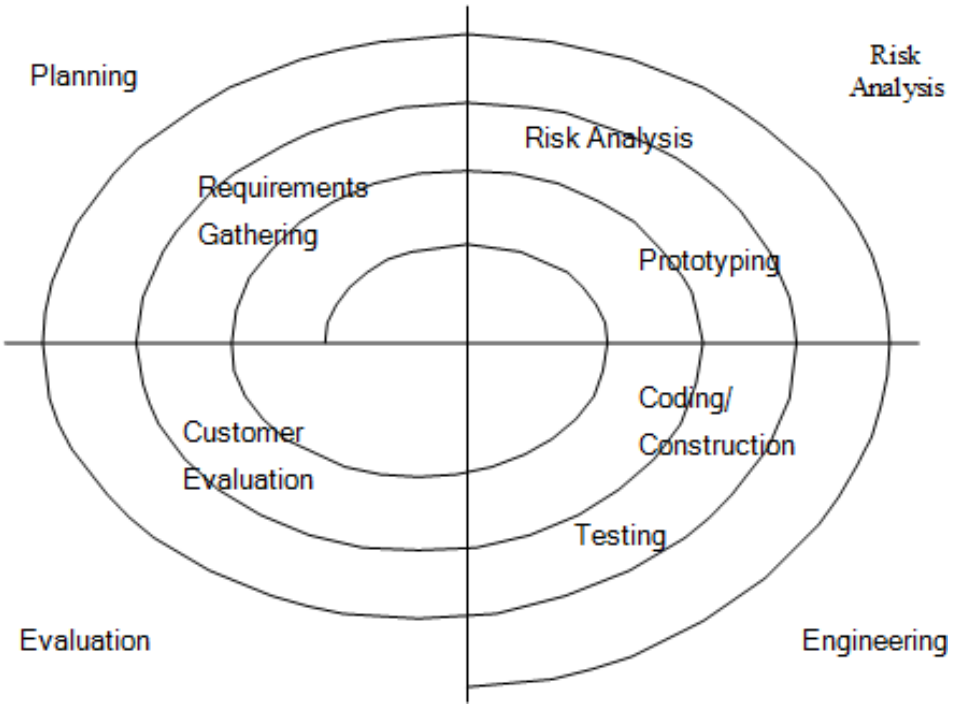
In order to solve this problem, we will use a real life hotel booking dataset to create a customer segmentation analysis in order to gain insights about the customers (and hopefully reasons why they cancel their reservation). We will then build a classification model (including the newly created customer clusters) to predict whether or not a booking will be canceled with the highest accuracy possible.

This model will allow hotels to predict if a new booking will be canceled or not, manage their business accordingly, and increase their revenue.

2. System Development Methodology.

The methodology of software development is the method in managing project development. There are many models of the methodology are available such as Waterfall model model, Incremental model, RAD model, Agile model, Iterative model and Spiral model. However, it still needed to be considered by developer to decide which will be used in the project. The methodology model is useful to manage the project efficiently and able to help developer from getting any problem during time of development. Also, it helps to achieve the objective and scope of the projects. In order to build the project, it need to understand the stakeholder requirements.

Methodology provides a framework for undertaking the proposed DM modelling. The methodology is a system comprising steps that transform raw data into recognized data patterns to extract knowledge for users.



There are four phases that involve in the spiral model:

1) Planning phase

Phase where the requirement are collected and risk is assessed. This phase where the title of the project has been discussed with project supervisor. From that discussion, Heart Prediction System has been proposed. The requirement and risk was assessed after doing study on existing system and do literature review about another existing research.

2) Risk analysis Phase

Phase where the risk and alternative solution are identified. A prototype are created at the end this phase. If there is any risk during this phase, there will be suggestion about alternate solution.

3) Engineering phase

At this phase, a software are created and testing are done at the end this phase.

4) Evaluation phase

At this phase, the user do evaluation toward the software. It will be done after the system are presented and the user do test whether the system meet with their expectation and requirement or not. If there is any error, user can tell the problem about system.

System Requirements:

1.13 Tools

For application development, the following Software Requirements are: Operating System: Windows 7 or any Linux Debian Distro.
Language: R and Shiny
Tools: RStudio IDE, Microsoft Excel (Optional).

Technologies used: R, Unix, Shiny.

Operating System Network
Visio Studio

Github
Google Chrome

1.13.1 Software requirements:

Any OS with clients to access the internet Wi-Fi Internet or cellular Network

Create and design Data Flow and Context Diagram
Versioning Control
Medium to find reference to do system testing, display and run shinyApp.

1.13.2 Hardware Requirements

For application development, the following Software Requirements are: Processor: Intel or high
RAM: 1024 MB
Space on disk: minimum 100mb

For running the application:
Device: Any device that can access the internet Minimum space to execute: 20 MB

The effectiveness of the proposal is evaluated by conducting experiments with a cluster formed by 3 nodes with identical setting, configured with an Intel CORETM i7-4770

processor (3.40GHZ, 4 Cores, 8GB RAM, running Ubuntu 18.04 LTS with 64-bit Linux 4.31.0 kernel)

1.14 Budget.

The budget of completion for developing the heart disease prediction system will require various software and hardware devices. The application is averagely expensive to build but if happens to be as successful as the developer sees it to be it will bring forth enough profit to cover the costs undergone.

3. Dataset Description

It contains information about bookings from two hotels in Portugal for the period from July 2015 to August 2017. One of the hotels is a resort hotel and the other is a city hotel. There are 119 390 observations in total (each of them describes one reservation). The key is that about 37% of them were cancelled which is a pretty large number. The dataset provides 31 attributes that can be useful in prediction, but we have selected only the 17 most promising, removing the target leakage or redundant features. The features used in this dataset are as following:

Variable Description

adr Average Daily Rate, calculated by dividing the sum of all lodging transactions by the total number of staying nights

adults Number of adults

agent ID of the travel agency that made the booking

arrival_date_week_number Week number of the arrival date

booking_changes Number of changes/amendments made to the booking from the moment the booking was entered on the Property Management System (PMS) until the moment of check-in or cancellation

country Country of origin

customer_type Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; Group - when the booking is associated to a group; Transient - when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party - when the booking is transient but is associated to at least another transient booking

hotel Type of hotel

lead_time Number of days that elapsed between the entering date of the booking into the PMS and the arrival date

market_segment Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"

previous_bookings_not_canceled Number of previous bookings not canceled by the customer prior to the current booking

previous_cancellations Number of previous bookings that were canceled by the customer prior to the current booking

required_car_parking_spaces Number of car parking spaces required by the customer

reserved_room_type Code of room type reserved. Code is presented instead of designation for anonymity reasons

stays_in_week_nights Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

stays_in_weekend_nights Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

total_of_special_requests Number of special requests made by the customer (e.g. twin bed or high floor)

PYTHON:

NUMPY:

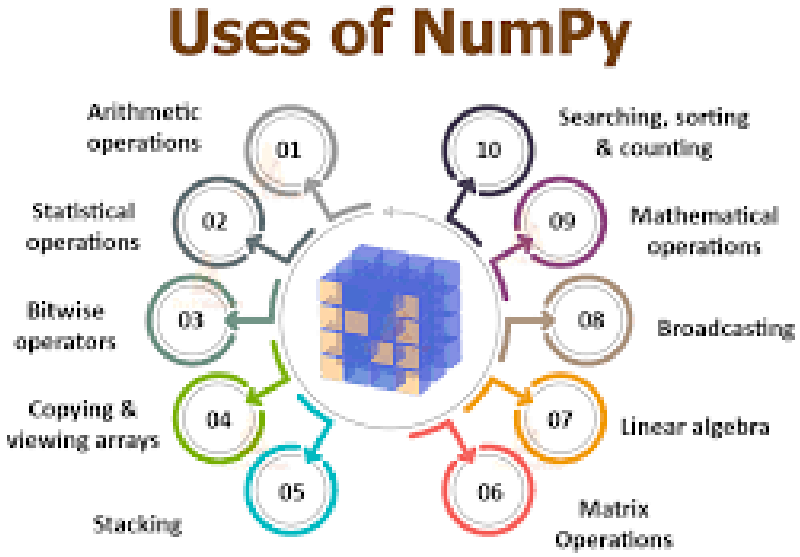
NumPy is the fundamental package for scientific computing in Python.It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

At the core of the NumPy package, is the ndarray object. This encapsulates n-dimensional arrays of homogeneous data types, with many operations being performed in compiled code for performance. There are several important differences between NumPy arrays and thestandard Python sequences:

- NumPy arrays have a fixed size at creation, unlike Python lists(which can grow dynamically). Changing the size of
  an ndarray will create a new array and delete the original.
- The elements in a NumPy array are all required to be of the same data type, and thus will be the same size in memory. The exception: one can have arrays of (Python, including NumPy) objects, thereby allowing for arrays of different sized elements.
- NumPy arrays facilitate advanced mathematical and other typesof operations on large numbers of data. Typically, such

operations are executed more efficiently and with less codethan is possible using Python’s built-in sequences.

- A growing plethora of scientific and mathematical Python- based packages are using NumPy arrays; though these typicallysupport Python-sequence input, they convert such input to NumPy arrays prior to processing, and they often output NumPy arrays. In other words, in order to efficiently use much (perhaps even most) of today’s scientific/mathematical Python-based software, just knowing how to use Python’s built-in sequence types is insufficient - one also needs to know how to use NumPy arrays.



PANDAS:

The Pandas library is one of the most important and popular tools for Python data scientists and analysts, as it is the backbone of many data projects. Pandas is an open source Python package for data cleaning and data manipulation. It provides extended, flexible data structures tohold different types of labelled and relational data. On top of that, it isactually quite easy to install and use.

Pandas is often used in conjunction with other data science Python libraries. In fact, Pandas is built on the NumPy package, so a lot of thestructure between them is similar. Pandas is also used in SciPy for

statistical analysis or with Matplotlib for plotting functions. Pandas canbe used on its own with a text editor or with Jupyter Notebooks, the ideal environment for more complex data modelling. Pandas is available for most versions of Python, including Python3.

Think of Pandas as the home for your data where you can clean, analyse, and transform your data, all in one place. Pandas is essentiallya more powerful replacement for Excel. Using Pandas, you can do things like:

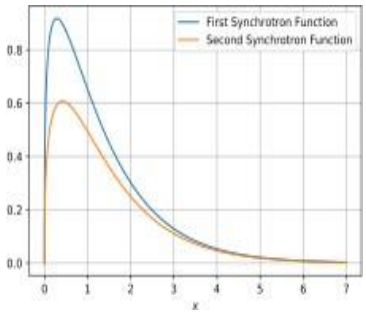
- Easily calculate statistics about data such as finding theaverage, distribution, and median of columns
- Use data visualization tools, such as Matplotlib, to easily createplot bars, histograms, and more
- Clean your data by filtering columns by particular criteria oreasily removing values
- Manipulate your data flexibly using operations like merging,joining, reshaping, and more
- Read, write, and store your clean data as a database, txt file,or CSV file

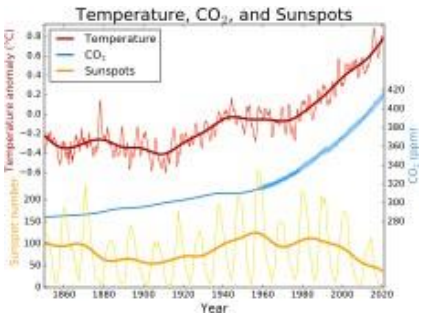
MATPLOTLIB:

Matplotlib is a plotting library for the Python programming language andits numerical mathematics extension NumPy. It provides an object- oriented API for embedding plots into applications using general- purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. There is also

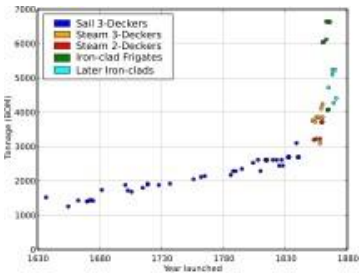
a procedural "pylab" interface based on a state machine (like OpenGL),designed to closely resemble that of MATLAB, though its use is discouraged.[3] SciPy makes use of Matplotlib.

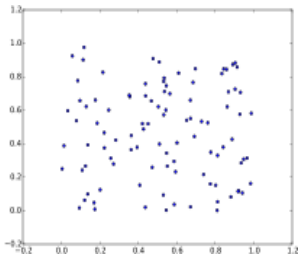
There are numerous plots that can be performed using matplotlib. During visualization of the dataset, matplotlib acts as a great medium to help users identify pattern and analyse datasets as much as possible. There are some instances where different plots play a more significant role thanas compared to the other ones. So, here are some plots that come in handy while performing exploratory data analysis:



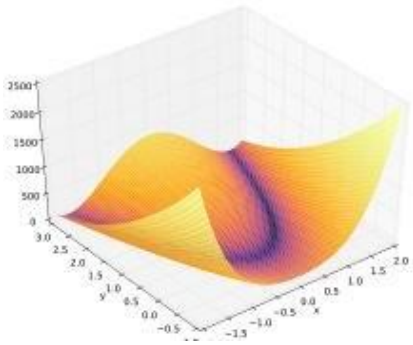


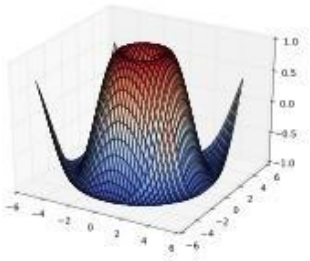
LINE PLOTS



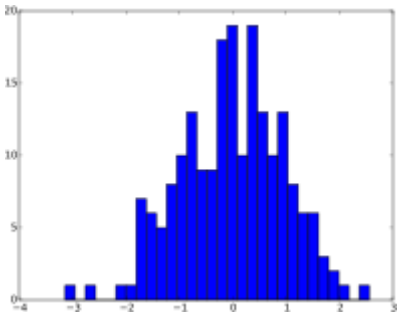


SCATTER PLOTS

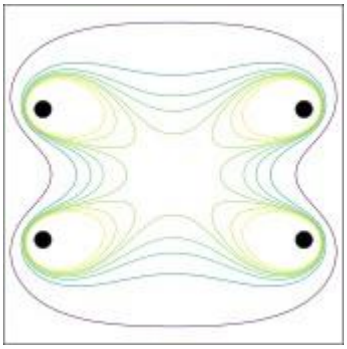

2500
2000
1500
1000
500
0
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
y
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
x


1.0
0.5
0.0
−0.5
−1.0
−6
−4
−2
0
2
4
6

3D PLOTS



HISTOGRAM



CONTOUR PLOT

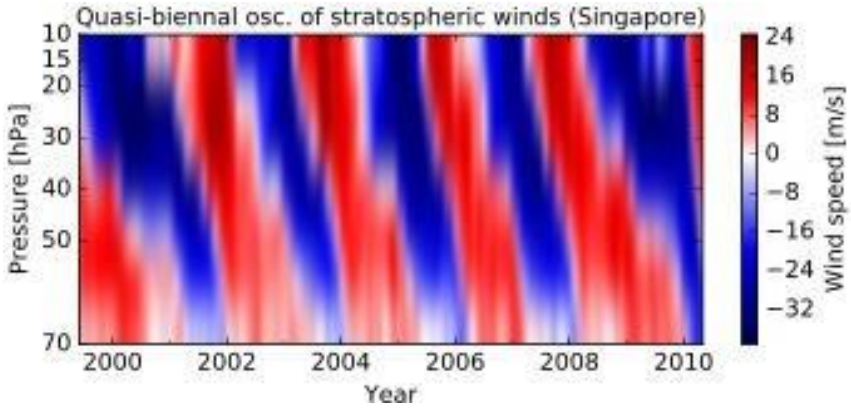


IMAGE PLOT

MACHINE LEARNING ALGORITHMS:

Linear Regression:

A Linear regression may be defined as the statistical model that analyses the linear relationship between a dependent variable with given set of independent variables. Linear relationship between variables means that when the value of one or more independent variables will change (increase or decrease), the value of dependent variable will also change accordingly (increase or decrease).

Mathematically the relationship can be represented with the help of following equation −

$Y = mX + b$

Here, Y is the dependent variable we are trying to predict

X is the dependent variable we are using to make predictions.

m is the slop of the regression line which represents the effect X has on Y

b is a constant, known as the Y-intercept. If X = 0,Y would be equal to b.

Positive Linear Relationship:

A linear relationship will be called positive if both independent and dependent variable increases. It can be understood with the help of following graph −

Negative Linear relationship:

A linear relationship will be called positive if independent increases and dependent variable decreases.

Linear regression types −

- Simple Linear Regression
- Multiple Linear Regression Simple Linear Regression (SLR)

It is the most basic version of linear regression which predicts a response using a single feature. The assumption in SLR is that the two variables are linearly related.

Multiple Linear Regression (MLR)

It is the extension of simple linear regression that predicts a response using two or more features.

Finding the best fit line: When working with linear regression, our main goal is to findthe best fit line that means the error between predicted values and actual values should be minimized. The best fit line will have the least error. The different values for weights or the coefficient of lines (a0, a1) gives a different line of regression, so we need to calculate the best values for a0 and a1 to find the best fit line, so to calculate this we use cost function.

The Goodness of fit determines how the line of regression fits the set of observations. Theprocess of finding the best model out of various models is called optimization. It can be achieved by R-square method.

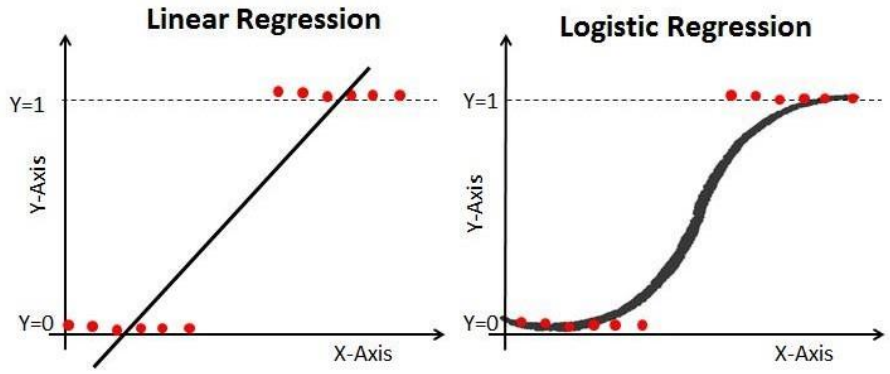
Below is an example of linear regression. Here I have identified the best fit line having linearequation y=0.2811x+13.9. Now using this equation, we can find the weight, knowing the height of a person.

Logistic Regression:

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of targetor dependent variable is dichotomous, which means there would be only two possible classes.

In simple words, the dependent variable is binary in nature having datacoded as either 1 (stands for success/yes) or 0 (stands for failure/no).

Mathematically, a logistic regression model predicts P(Y=1) as a function of X. It is one of the simplest ML algorithms that can be usedfor various classification problems such as spam detection, Diabetes prediction, cancer detection etc.



Types of Logistic Regression:

Binary or Binomial

In such a kind of classification, a dependent variable will have only twopossible types either 1 and 0

Multinomial:

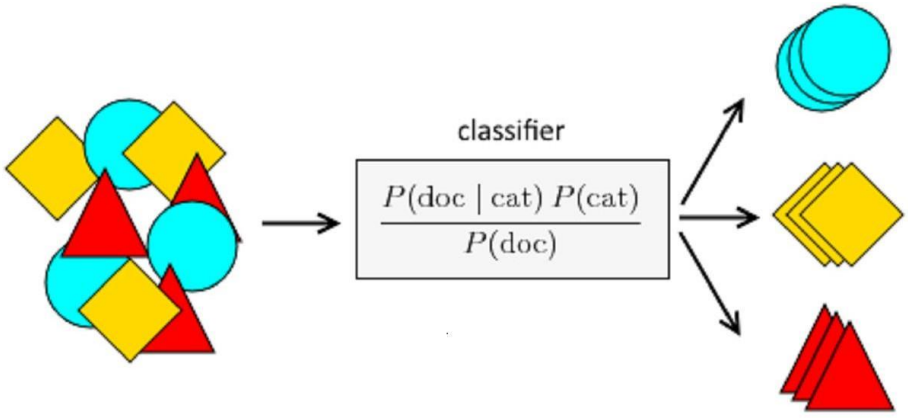
In such a kind of classification, dependent variable can have 3 or morepossible unordered types or the types having no quantitative significance.

Ordinal

In such a kind of classification, dependent variable can have 3 or morepossible ordered types or the types having a quantitative significance.

Naïve Bayes:

Naïve Bayes algorithms is a classification technique based on applying Bayes' theorem with a strong assumption that all the predictors are independent to each other. In simple words, the assumption is that the presence of a feature in a class is independent to the presence of any other feature in the same class. For example, a phone may be considered as smart if it is having touch screen, internet facility, good camera etc. Though all these features are dependent on each other, they contribute independently to the probability of that the phone is a smart phone.



In Bayesian classification, the main interest is to find the posterior probabilities i.e. the probability of a label given some observed features, P (L | features). With the help of Bayes theorem, we can express this in quantitative form as follows −

$$P(L|features)=P(L)P(features|L)P(features)P(L|features)$$

$$=P(L)P(features|L)P(features)$$

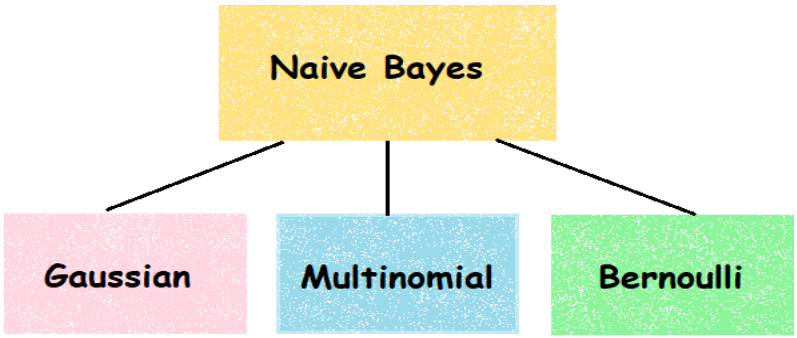
Here, $P(L \mid features)$ is the posterior probability of class.

$P(L)$ is the prior probability of class.

$P(features \mid L)$ is the likelihood which is the probability of predictor given class.

$P(features)$ is the prior probability of predictor.

TYPES OF NAÏVE BAYES:



Gaussian Naïve Bayes

It is the simplest Naïve Bayes classifier having the assumption that thedata from each label is drawn from a simple Gaussian distribution.

Multinomial Naïve Bayes

Another useful Naïve Bayes classifier is Multinomial Naïve Bayes in which the features are assumed to be drawn from a simple Multinomialdistribution. Such kind of Naïve Bayes are most appropriate for the features that represents discrete counts.

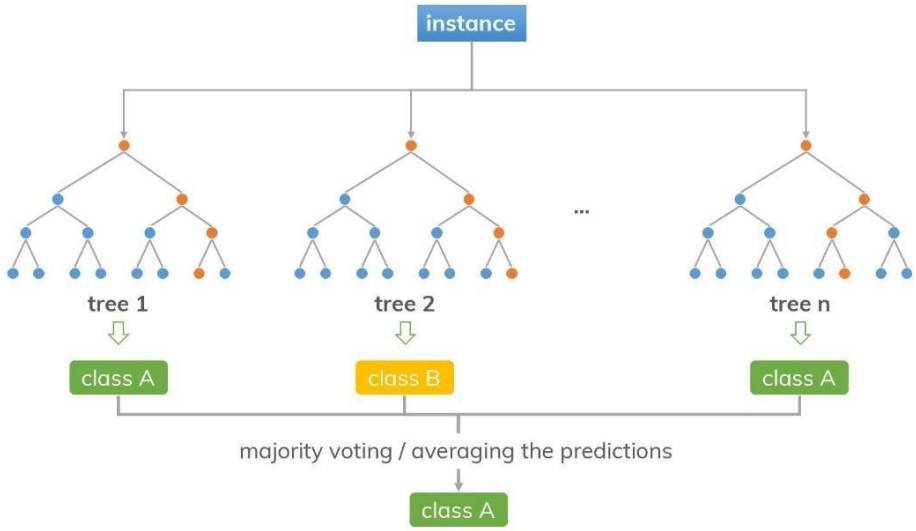
Bernoulli Naïve Bayes

Another important model is Bernoulli Naïve Bayes in which features are assumed to be binary (0s and 1s). Text classification with 'bag of words' model can be an application of Bernoulli Naïve Bayes.

Random forest:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of



that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

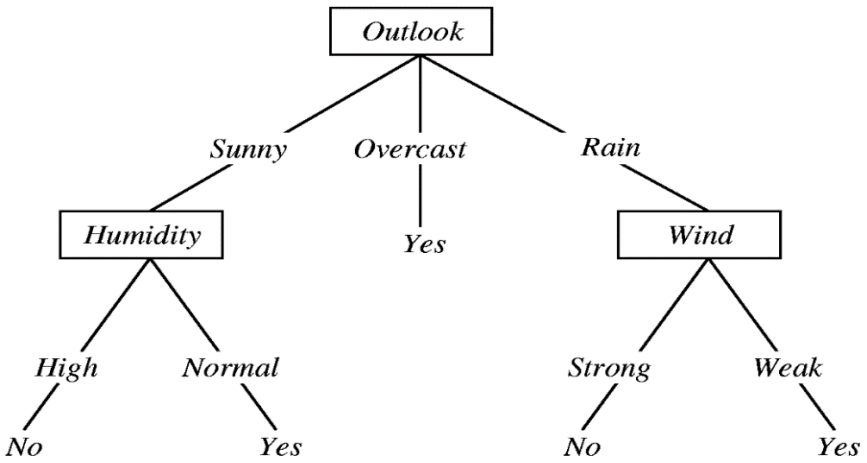
Applications of Random Forest:

There are mainly four sectors where Random Forest mostly used:

1. Banking: Banking sector mostly uses this algorithm for the identification of loan risk.
2. Medicine: With the help of this algorithm, disease trends and risks of the disease can be identified.
3. Land Use: We can identify the areas of similar land use by this algorithm.
4. Marketing: Marketing trends can be identified using this algorithm.

Decision Tree:

A decision tree is a flowchart-like structure in which each internal node represents a test on a feature (e.g. whether a coin flip comes up heads or tails) , each leaf node represents a class label (decision taken after computing all features) and branches represent conjunctions of features that lead to those class labels. The paths from root to leaf represent classification rules.



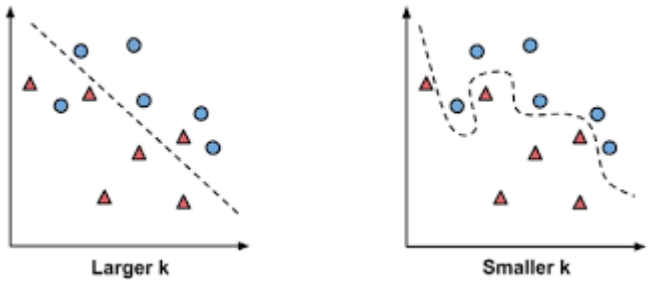
Decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are non-parametric supervised machine learning method used for both classification and regression tasks. the condition refers to a range. Each leaf is assigned to one class representing the most appropriate target value. Alternatively, the leaf may hold a probability vector indicating the probability of the target attribute having a certain value. Instances are classified by navigating them from the root of the tree down to a leaf, according to the outcome of the tests along the path. Figure 9.1 describes a decision tree that reasons whether or not a potential customer will respond to a direct mailing. Internal nodes are represented as circles, whereas leaves are denoted as triangles. Note that this decision tree incorporates both nominal and numeric attributes. Given this classifier, the analyst can predict the response of a potential customer (by sorting it down the tree) and understand the behavioral characteristics of the entire potential customers population regarding direct mailing. Each node is labeled with the attribute it tests, and its branches are labeled with its corresponding value

kNN:

Machine learning techniques have been widely used in many scientific fields, but its use in medical literature is limited partly because of technical difficulties. k-nearest neighbors (KNN) is a simple method of machine learning. The article introduces some basic ideas underlying the KNN algorithm, and then focuses on how to perform KNN modeling with R. The dataset should be prepared before running the KNN function in R. After prediction of outcome with KNN algorithm, the diagnostic performance of the model should be checked. Average accuracy is the mostly widely used statistic to reflect the KNN algorithm. Factors such as k value, distance calculation and choice of appropriate predictors all have significant impact on the model performance.

The impact of selecting a smaller or larger K value on the model

- Larger K value**:** The case of underfitting occurs when the value of k is increased. In this case, the model would be unable to correctly learn on the training data.
- Smaller k value**:** The condition of overfitting occurs when the value of k is smaller. The model will capture all of the training data, including noise. The model will perform poorly for the test data in this scenario.



Here are some things to keep in mind:

1. As we decrease the value of K to 1, our predictions become less stable. Just think for a minute, imagine K=1 and we have a query point surrounded by several reds and one

green (I'm thinking about the top left corner of the colored plot above), but the green is the single nearest neighbor. Reasonably, we would think the query point is most likely red, but because K=1, KNN incorrectly predicts that the query point is green.

2. Inversely, as we increase the value of K, our predictions become more stable due to majority voting / averaging, and thus, more likely to make more accurate predictions (up to a certain point). Eventually, we begin to witness an increasing number of errors. It is at this point we know we have pushed the value of K too far.

3. In cases where we are taking a majority vote (e.g. picking the mode in a classification problem) among labels, we usually make K an odd number to have a tiebreaker.

6. COMPARE ALL MODEL

```
LogisticRegression
[[15002  4903]
 [ 3637  6261]]
0.7134516659396705


Naive Bayes
[[ 8547  1253]
 [10092  9911]]
0.6193336241317988


RandomForest
[[18523  1250]
 [  116  9914]]
0.9541656880179847


Decision Tree
[[17802   829]
 [  837 10335]]
0.9440995872898702



KNN
[[18484  1440]
 [  155  9724]]
0.9464818977955239
```

As we can see that the dataset that have is best suited for KNN, Random Forest or Decision Tree, particularly in that order.

7. CONCLUSION AND FUTURE WORK

After going through this project, we can clearly see that machine learning models can be easily created with the use of multiple machine learning algorithms. With the use of all the machine learning algorithms, we can even check which algorithm suits our dataset the best. We check their accuracies; we can perform further changes by putting in extra effort in data wrangling and cleaning to gain a much better accuracy and deploy the model so as to make sure that the users get to use this to the best of its capability.

This project is just the beginning of the topic. There are many other fields where this project can be further enhanced and used to perform multiple other tasks. For e.g., we can easily create a standalone application to further expedite the prediction process, where in the application would easily detect which customer are most likely to drop their reservations. By pursuing this, there will be no human interference at all. The booking would be easily checked on its own and the user would directly get notified whether the customer is likely to cancel the reservation or not.

With a little change, this can further be used in different areas like the restaurant system where they detect a booked table would get cancelled or not. Same thing goes banquet halls. This can easily be used to predict whether a particular event booking would be performed or not. All in all, there are numerous such cases where this would be quite convenient.

Bibliography

1. https://machinelearningmastery.com
2. https://ai.google
3. https://towardsdatascience.com
4. https://youtube.com
5. https://www.tensorflow.org
6. https://www.javatpoint.com
7. https://www.stackoverflow.com
8. https://www.quora.com
9. https://www.analyticsvidhya.com