Divyesh Sethiya

Data Science Certification

2021-2022

[Linkedin](#)

SELF PROJECT

Predictive Modelling for Default of Credit Card User

Based on the data we have to develop a classification model to find whether a bank customer will default their credit card loan payment for the next month or not. Considered Various Machine Learning Classification Models and metrics to devise a Robust Model that can classify whether the Person is going to be Default or not.

Skills and Tools

PYTHON , EDA , STATISTICS , MACHINE LEARNING

COURSE PROJECTS

Apr'22

Machine Learning based Clustering and PCA Problem

Course Unsupervised Learning

Based on given data of visitors browsing for online shopping, build different clusters to know whether person is only browsing and visiting multiples pages or also generating revenue for the shopping websites as well. Analyse and compare the clusters formed with the existing Revenue Column.

Skills and Tools

Machine Learning, Unsupervised Learning, K-means clustering, Hierarchical Clustering, Principal Component Analysis

Apr'22

Machine Learning based Disease Prediction Model

Course Supervised Learning - Classification

Build the binomial machine learning model to predict the Parkinson disease. The objective of this model is to classify the patients with the Parkinson disease. Perform the required pre-processing steps prior to model building and evaluate the model performance through proper tuning process.

Skills and Tools

Machine Learning, Supervised Learning, Linear and Non Linear Classification model, Hyper parameter, Model Metrics

Mar'22

Machine Learning model building for continuous variable prediction

Course Supervised Learning - Regression

Build a Linear Machine Learning model to understand the relationship between the population of US cities in the year 1920 and 1930. Evaluate the model performance with appropriate measures. Perform all the required graphical and quantitative exploratory data analysis prior to model building. The dataset has 49 rows and 2 columns. The measurements are the population (in 1000's) of 49 U.S. cities in 1920 and 1930. The 49 cities are a random sample taken from the 196 largest cities in U.S

Skills and Tools

Machine Learning, Supervised Learning, Linear Regression

Feb'22

Descriptive Statistics on real world problem

Course Statistics for Machine Learning

The first problem statement focuses on exploratory data analysis on a la liga cup (a football tournament). Learners need to analyze the data to get useful insights on top teams, winning distribution etc. The second problem statement helps in understanding and applying confidence intervals to a healthcare related problem. The third problem statement will make learners apply statistical techniques like design experiment and hypothesis testing on designing golf data.

Skills and Tools

Descriptive and inferential statistics, Hypothesis testing, confidence interval, Distribution.

Feb'22

Sports Analytics

Course Exploratory data analysis

A new football club named 'Brussels United FC' has just been inaugurated & the team is looking to hire players for their roster. Management wants to make such decisions using data based approach. Player data for all teams has been acquired from FIFA. The data contains details for over 18,000 players playing in various football clubs in Europe. It contains information on age, skill rating, wages and player value, etc. in the files named fifa.csv which is data file and fifa_variable_information.cs

Skills and Tools

ITP, NPV and EDA

Jan'22

Online Shopping Process Analytics

Course Database Management Systems - II

To learn and understand the purchasing behavior of online customers, the data is confined to 'Sales and Delivery' and is provided for the period of last decade. This will be useful in · Analyzing seasonality & business patterns · Product analysis · User analysis . order placed . orders delivered

Skills and Tools

MySQL queries

Dec'21

ITP/NPV

Course Introduction to Python Programming

Perform the data analysis to get the different insights using the data collected from the IPL matches. The dataset contains 2 files 'deliveries.csv' and 'matches.csv'. matches.csv contains details related to the match such as location, contesting teams, umpires, results, etc. deliveries.csv contains the information about each delivery bowled in all the IPL matches line runs scored, batsman, bowler, non-striker, etc.,

Skills and Tools

Python, Numpy, Pandas