Projects Completed in Data Science

Projects using Python and R

Project 1 Innodatatics, Hyderabad (Aug 2020 to Nov 2020)

Project Title : Identifying Risk factors influencing Diabetes Type – Classification Model.

Domain : Health Care

Roles and Responsibilities : **Junior Data Analyst- Team Leader**

Problem Statement

Collection and analysis of data received from the diagnostic department which had recorded data on a few regular patients who visit hospital more often. The dataset consisted of data recorded from 2018 to 2019. The dataset focuses on patients suffering from "Type 1 diabetes", "Type 2 diabetes", and a few other diseases. The aim is to predict and cluster the patients as per the types of diabetes they suffer and admit them in the proper ward. Building a machine learning model to group them accordingly to clear up the confusion that is being created in the hospital. The purpose of this project is to predict the ability of machine learning for classifying the types of Diabetes Mellitus (DM) using factors that determine diabetic types in susceptible individual, thus helps health care providers in different settings to find out the disease in their earliest phase which in turn enhance good prognosis which will bring good business benefits.

Technology Stack and Algorithms used :

Python : Machine Learning

Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow

ML Algorithms Used : In this project, Unsupervised learning namely K Means clustering and Gradient Boosting Regression algorithm are used for obtaining high accuracy percentage. Also Supervised learning algorithms namely Neural Network, LR, LDA, KNN, CART, NB, and SVM, is used to effectively verify Model accuracy of the models.

Steps Used: Descriptive Analysis and Predictive Analysis and for all steps of variable description, exploratory data analysis, data processing and preprocessing, feature selection, model building, and, evaluation of models are done using Python and R.

Project 2- Innodatatics , Hyderabad (Dec 2020 to May 2021)

Project Title:

Data Exploration and Forecasting on Patient Count – Regression Model

Domain : Health Care

Roles and Responsibilities : **Junior Data Analyst- Team Leader**

Problem Statement

The purpose of the project is to predict the ability of Machine Learning (ML) techniques to analyze the data pertaining to Out-patients visiting different hospitals with respect to year-wise, age-wise, stage-wise, specialty-wise,time-bound. Our approach involves descriptive analysis and processing of data, building machine learning models and evaluating their accuracies, identifying challenges and constraints in the execution of the projectwhich will bring good business benefits.

Technology Stack and Algorithms used :

Python : Machine Learning

Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow

ML Algorithms Used : Unsupervised Learning Algorithms namely Arima() Model, Auto Regressor() Model,Stationarity Analysis.

Steps Used: Descriptive Analysis and Predictive Analysis and Time Series Analysis for all steps of variable description, exploratory data analysis, data processing and preprocessing, feature selection, model building, and evaluation of models are done using Python and R.

Libraries used: Pandas, Scikit-learn, Numpy, Keras, Tensor flow

---

Project 3 Innodatatics , Hyderabad (July 2021 to Sept 2021)

Project Title :Infant Mortality Rate Analysis –Regression Model

Roles and Responsibilities : **Junior Data Analyst- Team Leader**

Problem Statement

To find the proportion of infant death registrations according to their respective GEO information To visualize each age category at the time of death of an infant to understand their death rate ,To predict/forecast the total no. of deaths for every age group listed in the data for the next 5 years.

Technology Stack and Algorithms used :

Python : Machine Learning

Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow

ML Algorithms Used : Unsupervised Learning Algorithms namely Arima() Model, Auto Regressor() Model,Stationarity Analysis.

Steps Used: Descriptive Analysis and Predictive Analysis and Time Series Analysis for all steps of variable description, exploratory data analysis, data processing and preprocessing, feature selection, model building, and,evaluation of models are done using Python and R.

Libraries used: Pandas, Scikit-learn, Numpy, Keras, Tensor flow,

---

Project 4 Innodatatics , Hyderabad (Oct 2021 to Dec 2021)

Project Title :Survival Analysis – Unemployment Index–Classification Model

Domain : General -Social

Roles and Responsibilities: **Junior Data Analyst**

Problem Statement

- To describe the unemployment index of survival group with respect to spell using Supervised Machine Learning method.
- To describe the unemployment index of unemployment group with respect to spell using Supervised Machine Learning method.
- To classify the unemployment index of survival and unemployment group with respect to timeline estimate using Supervised Machine Learning method.
- To apply suitable machine learning model for classifying the unemployment index with respect to spell and event
- To create awareness of unemployment rate so as to suggest ways and means to minimize unemployment rate

Constraints:

The dataset has very few variables. Limited open source and cross platform to study the dataset. A possibility of model under fitting due to low variance

Project Pipeline : Data Understanding,Data Cleaning ,Machine Learning Model Engineering

Model Selection-Kaplan-Meier Estimator ,Model Testing and Evaluation,Model Deployment

Monitoring and Maintenance

.

Technical Stacks and Algorithms used in this Project

Languages: Python, html, cloud, sql, heroku, R, visual studio, git bash

AI/ML: Pytorch, skikit learn,

Libraries: lifelines, pandas, numpy, skikit learn

Database: Jupyter Online

Warehouse: Jupyter classic notebook

ETL: Python, Jupyter

Visualizations: Plotly()

Tracking & SC: Github

Deployment Strategy:

Web Based- HTML applications

Enterprise Based- AI algorithm/ Kaplan Meier Fitter Model

Project 5 - Intellipaat Bengaluru

Domain: Banking

Project Description Loan Eligibility Prediction – Classification Model

Technology Stack and Algorithms used :

Python : Machine Learning

Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow
ML Algorithms Used: In this project, Unsupervised learning namely K Means clustering and Gradient Boosting Regression algorithm are used for obtaining high accuracy percentage. Also Supervised learning algorithms namely Neural Network, LR, LDA, KNN, CART, NB, and SVM, is used to effectively verify Model accuracy of the models.

Projects 6-Intellipaat ,Bengaluru using Tableau Desktop
Project : Covid-19 mortality rates.
Domain: Healthcare
Roles and Reaponsibilities Project Intern in Data Science Architect Masters Program
Project Description:
Based on the csv.files for global mortality rates, we have analyzed and developed a dashboard to understand the covid-19 global cases. Using filters, parameters and actions wherever possible to make the dashboard interactive.Comparing the global confirmed vs. death cases in a world map using pie charts.Creating a parameter for percentile comparison between countries based on confirmed or death cases. Comparing the country wise cases using logarithmic axes. Dashboard displaying both log axis chart and a default axis chart in an alternate interactive way. New cases per day in China and India – compared in a date wise chart.Which day has the highest new death cases?
Average WHO region wise cumulative cases visualized using a funnel chart. Dashboard created with a drop down menu to view the WHO region wise data using a bar chart, line chart or a map as per user's requirement.

Project 7 – Intellipaat ,Bengaluru using Microsoft SQL
Project Domain: Sales Analysis
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Project Description:
Querying a large relational database using Adventure work database with SQL server with respect to.

- details from the personal table including email ID, phone number, and phone number type
- details of the sales header order
- details of the sales details order made in a month
- total sales made in a month
- total sales made in the year by month order by increasing sales
- total sales made to the customers

Project 8 - Intellipaat ,Bengaluru
Project Domain: Travel
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Big Data Hadoop Spark
Project Description:
Analysis of Taxi App pertaining to number of trips,no of passengers travelled, total number of kilometres travelled, rate per trip and revenue generated.

Project 9- Intellipaat ,Bengaluru
Project Domain: E **commerce website**
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Big Data Hadoop
Project Description
Analysis of the public review of the products on the Social media
Java, eclipse, neon, javascript, java compiler, apache

Project 10- Intellipaat ,Bengaluru
Domain :Social Media

Project Domain: E **commerce website**
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Project Description
Analysis of Cosmetic Products on Social Media Platform with
respect to likes, dislikes
Technology used :Spark

Project 11Intellipaat , Bengaluru
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Project Description
Title : Binary Classification on Customer Churning using Keras
Analysis of reasond why customers are switching to other competitors
Domain : Telecom Industry
Technology used :Artificial Intelligence keras

Project no 12 – Intellipaat , Bengaluru
Roles and Reaponsibilities : Project Intern Data Science Madder's Architect Program
Project Domain :Bank Customer Data
Project Description
The objective of this project is to understand the UK bank customer data
Based on the .csv file provided analysis was done and developed a dashboard to understand the customer data of a UK bank. Use filters, parameters and actions wherever possible to make the dashboard interactive
Technology used :Tableau desktop
Interactive dashboard wasused to understand the data. Multiple features, various charts suitable to the respective problem statements are to be used at points where chart type is not specified

---

Project No 13 – Intellipaat , Bengaluru
Roles and Reaponsibilities : Project Intern Data Science Madder's Architect Program
Project Domain :E commerce
Project Description :
Building an AI-based Chatbot using IBM Watson LAB
Domain :E Commerce
Technology used : Artificial Intelligence & Deep Learning with Tensorflow

---

Project No 14 – Intellipaat , Bengaluru
Roles and Reaponsibilities : Project Intern Data Science Madder's Architect Program
Project Domain :Analytics
Project Description
K-Means cluster analysis on Iris dataset to predict the class of a flower using its petal's Dimensions .Using the famous Iris dataset, predict the
class of a flower Perform K-Means cluster analysis
Technology used :SAS

International Data Science Kaggle Project Competitions
Kaggle Status: Notebook Expert / Competitions Contributor
Kaggle Profile : https://www.kaggle.com/gopalkk1
Winner of 10 bronze medals

Participated in 35 of Data Science Competitions

Projects done in training program with IIT Kanpur – Jan 2023- Sept 2023

DATA SCIENCE PROJECTS DONE DURING DATA SCIENCE TRAINING PROGRAMS–Edvancer-IIT Kanpur

Machine Learning Projects- Python

Consumer Services- Consumer Complaints Resolutions

- ✓ BFSI Marketing- Understanding Customer Preferences in Insurance Sector
- ✓ Pharma Public Safety- Counterfeit Medicines- Prediction of Sales
- ✓ Manufacturing- Predict Hazard Ratings for a Maintenance Project
- ✓ Real Estate-Flag Junk Property Listings
- ✓ Health Care- To Predict No-shows given the appointment details

Artificial Intelligence Projects:

- ✓ Multiclass Multilabel prediction For stack overflow- Given text for Questions , predict tags associated with them
- ✓ Music Genre Identification- Given audio files for songs , identify which genre they fall in
- ✓ Spam filter for Quora questions
- ✓ Distracted Driver Multi Action Classification Page-Classification of the various distractions of a driver
- ✓ Image Captioning Page- Uploading photos and pixel analysis

---