PAVAN KUMAR DUMALE

PROFILE

Strong Data science knowledge with a passion to solve real-world business challenges using data analytics. Proficient in deploying complex machine learning and statistical modeling algorithms / techniques for identifying patterns and extracting valuable insights for key stakeholders and organizational leadership.

TECTECHINCAL SKILLS

Packages :
Scikit-learn, Pandas, Numpy, plot.ly, Matplotlib, Seaborn

Software :
Python, R Language, Tableau, SQL, Excel

Statistics and ML:
Logistic regression, Linear Regression, SVM, KNN, Decision Tree, Ensemble Learning algorithms, K-means clustering

Certifications:

- Certificate of Excellence in business Analytics
- Certificate of Excellence in Machine learning

CONTACT

+91 8096107275

Email:
pavankumardumale@gmail.com

WEBSITE:
linkedin.com/in/pavan-kumar-15211a1a4

EDUCATION AND QUALIFICATION

QUALIFICATION	UNIVERSITY / COLLEGE	DURATION	AGGREGATE
B.Tech – Mechanical Engineering	Jawaharlal Nehru Technological University, Hyderabad	2014-2018	65.75%
Intermediate	Srichaitanya jr. College, Hyderabad	2012-2014	80.7%
SSC	Vamshi High School, Bodhan	2012	9.3 GPA

EXPERIENCE AND AREAS OF STRENGTH

DATA ANALYTICS :

- Proficient in Understanding the Business problems Applying Descriptive, Diagnostics, Predictive, Prescriptive Analysis to give best possible solutions.

FEATURE ENGINEERING:

- Data Gathering through different sources SQL, API, Kaggle , on premise etc.
- Advanced and keen Handling of raw data with various techniques, Treatment of Null values, missing values, outliers, Categorical Features, Text data etc.

DATA VISUALIZATION:

- Steering and rapid model creation in Python using Scikit Learn, Pandas, Numpy, Matplotlib , Seaborn for Data visualization.
- Expert in Understanding the valuable Insights, patterns, trends hidden in the data, and relationship between features.

STATISTICS :

- Exploratory data analysis, Data Quality, Hypothesis Testing, Design of Experiments, ANOVA, Regression Models, Functional Models, Classification Models, Bagging, Boosting, Forecasting (ARIMA) Models , Model Assessing, Model validation, Clustered Models.

MACHINE LEARNING:

- Hands on experience in Building various Machine Learning Models according to the data and business problem.
- Worked on Linear Regression, Logistic Regression, SVC, SVR, KNN, Naïve Bias, Decision Tree, Ensemble Learning algorithms Random Forest, AdaBoost, Gradient Boost, XGBoost, k-means clustering, Forecasting.
- Built APIs using the Flask for deploying the model on local host.

SQL SERVER:

- Interacting with databases, have good knowledge in using SQL queries.
- Hands on Experience in DQL, DDL, DCL, and DML commands

HOBBIES

- When I'm not in front of Computer like to play Outdoor games Cricket.
- I do stalk Tech blogs on Social Media.

Languages:

English, Hindi, Telugu, Marathi, Kannada

PROJECTS:

- Completed few Projects From the scratch Understanding the Business problem to Building Model, Understanding the Insights in the problem, model deployment, giving possible and valuable solutions.

Project Details:

Role : Business Analyst & Junior Data Scientist

Project : Worked on Pet Adoption Dataset from Hacker earth competition.

Problem Statement:
A leading pet adoption agency plans on creating a virtual-tour experience, showcasing all animals available in their shelter. We have been tasked to build a machine learning model that determines Pet type and breed of the animal based on its physical attributes and other factors.

- I have done exploratory data analysis, data summarization, and visualization to the complete dataset for better understanding that helped me to frame the predictive modelling problem.
- Applied Statistical Techniques for Problem Framing
- Using Summary statistics summarized distribution and understood relationships between each variable
- Cleaned the Data using Data corruption, Data errors and Data loss
- With the help of Data selection I have selected features which are having more impact on dependent variable
- Scaling, Encoding, Transforms are done in the part of Data preparation
- Selected Random Forest algorithms for prediction. As this Dataset has two predictions columns with multi class classification and imbalanced data.
- I chose Random Forest it is Ensemble Learning Technique and uses bootstrapped aggregation which leads to higher performance.
- Then created an API gateway to interact with model in local host.