Shahid **Farhan**

BE-COMPUTER SCIENCE ENGINEERING · VELAMMAL ENGINEERING COLLAGE · BS-DATA SCIENCE AND APPLICATIONS · IIT-MADRAS

☐ (+91) 8838628368 | ✉ shahidfarhan2207@gmail.com | ⊙ shahidfarhan22 | in shahidfarhanws

Summary

I am a data science enthusiast with a background in Computer Science and Statistics. I am also keenly interested in Statistical Analysis, Machine Learning and Data Visualization. My passion is to use data to solve problems and gain insights. I am always looking to stay up-to-date with the latest developments in the field and I am interested in pursuing a career in developing new data-driven products.

Education

Velammal Engineering Collage *Tamil Nadu, India*
B.E IN COMPUTER SCIENCE AND ENGINEERING | CGPA: $8.55/10$ *Aug. 2020 - present*

Indian Institute of Technology - Madras *Tamil Nadu, India*
BS IN DATA SCIENCE AND APPLICATIONS | CGPA: $6/10$ *July. 2020 - present*

St.Sebastian Matriculation Higher Secondary School *Tamil Nadu, India*
GRADE -12 | PERCENTAGE: 72.3 *2020*

St.Sebastian Matriculation Higher Secondary School *Tamil Nadu, India*
GRADE -10 | PERCENTAGE: 93.6 *2018*

Projects

Spam or Ham ⊙

ML-NLP,SPACY *Aug 2023*
- This project aims to classify whether a given email is spam or legitimate by employing techniques from natural language processing (NLP) for accurate and efficient analysis.
- Text preprocessing techniques-[Stemming, Lemmatization], Text Vectorization-[CountVectorizer from sklearn], Pipeline, Naive Bayes-Multinomial Naive Bayes.

Movie Genre Predictor ⊙

ML-NLP,SPACY,MULTINOMIAL NAIVE BAYES *July 2023*
- The primary goal of this project is to develop a robust and accurate system for automatically categorizing movies into distinct genres based on their plot summaries.
- Data Cleaning, Text preprocessing techniques -[Removing Stop words,Stemming, Lemmatization], Data Visualization, Text Vectorization - [CountVectorizer from sklearn], Model Building.

News Categorizer ⊙

ML-NLP,SPACY *June 2023*
- This project aims to automate the categorization process, ensuring that news content is efficiently classified into relevant topics through natural language processing. It offers a nuanced understanding of content, facilitating precise categorization.
- Cleaning the data, Handling imbalanced data, Categorical variable encoding, Preprocessing using en_core_web_sm -a small English pipeline trained on written web text, Naive Bayes-Multinomial Naive Bayes.

Diabetes Prediction Model ⊙

ML-REGRESSION, CLASSIFICATION, DECISION TREE, NAIVE BAYES *Jan 2023*
- Developed a machine learning model to predict the likelihood of diabetes in individuals based on relevant medical and demographic data which consists of various features such as age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level.
- Handled imbalanced data, Detected outliers, Categorical variable encoding, Feature Scaling.

SalesIQ

DATA ANALYSIS *Oct 2022*

- Unleashing Sales Insights Through Advanced Analytics
- Analyze the given data to improve sales and customer experience
- Analyze which type of product is selling the most
- Analyze during which period of the year the sales are highest

Skill

- Experienced in **Python** and Python **data science packages** like **Pandas, numpy, sklearn, regex**.
- **ML model building** and evaluation experience using **sklearn**.
- Able to perform **advanced text analysis** and **language modeling** using Natural Language Processing packages like **scapy**
- **Visualization skills** in python with **matplotlib, seaborn, plotly** etc. Ability to create customized plots using such tools.
- Active **Kaggle learner**
- Ability to **write effective, scalable and modular code**. Should be able to understand, test and **debug existing python** project modules quickly.
- Hands on experience with **DBMS software** such as **Postgresql** using **psycopg2** library in python.
- Knowledge on web development platforms such as **HTML,CSS,Flask**.
- Other skills such as **LaTeX, Touch typing(80 wpm)**.

Honors & Awards

2022	**Finalist**, Olympiad on Mathematics and Statistics - IITM BS Maths and Stat Academic Cell	*Certificate*
2022	**Completed** , Foundational Level in Programming and Data Science at - IIT Madras	*Certificate*
2022	**Published**, Paper titled - Subscribe with IOT for a Smart Soil Moisture Irrigation System - ICT4SD 2022	*Certificate*

Courseworks

Mathematics for Data Science I & II, Statistics for Data Science I & II, Programming in Python, Computational Thinking, Data Structures and Algorithms using Python, Database Management Systems, Big Data Analytics,Object Oriented Programming, Theroy of Computation, Design and Analysis of Algorithms,Discrete Mathematics, Resource Management Techniques, Artificial Intelligence

Extracurricular Activity

Paradox - 2k22 *Tamil Nadu, India*

CORE MEMBER *Jun. 2021 - present*

- Volunteered in Organizing the cultural fest of IITM-BS DataScience and Applications

Quizzing Club Indian Institute of Technology - Madras *Tamil Nadu, India*

CO-ORDINATOR *Jun. 2021 - present*

- Started quiz initiatives through Weekly quiz Challanges
- Organized Intra Houses quiz events

Teaching Volunteer *Tamil Nadu, India*

MEMBER *Mar. 2022*

- Volunteered in teaching Mathamatics for high school students
- Volunteerd in doing seminars periodically

Chess *Chennai, India*

PLAYER *present*

- Ranking - 1200
- Department player - Velammal Engineering Collage