Hemansh Anand

London, United Kingdom   ✉ anandhemansh@gmail.com   ☎ +44 07831254650   in hemanshanand

↖ Link to my Portolio Website   ⌨ Github: hemansanand   ⚑ Sponsorship Not Required

Summary

Certified full-stack Data Scientist having around 2 years of experience working on different projects in the Data Science domain. I have worked on projects involving gathering, exploring, and analyzing data to perform predictive, descriptive, and prescriptive analytics to provide actionable insights that act as the cornerstone for developing strategies for further product development. Highly skilled in Machine Learning and Data Visualization

Technical Skills

Languages & Aptitudes (Python, C, R , scikit-learn, TensorFlow, NLTK, libsvm, pyTorch, NumPy, Pandas, Unix)

Data Visualization (matplotlib, Seaborn, Plotly, Tableau) | **Regression** (Logistic,Linear,Random Forest)

Classification (K-NN,SVM,Random Forest,Naive Bayes) | **Neural Networks** (ANN,CNN, LSTM-RNN)

Database (MySQL,NoSQL, MongoDB) | **Web Development** (HTML, CSS) | **Hardware** (RaspberryPi, Arduino)

Frameworks & Cloud Services (Flask, Amazon Web Services (AWS), Google Cloud Platform(GCP), Azure, Heroku)

Certifications and Courses

Data Scientist Professional Certificate

Data Camp   Aug 2021

Machine Learning by Stanford University

Coursera   Jul 2020

Professional Experience

Junior Data Scientist   Sep 2019 – Jan 2021 | New Delhi, India

RFS India Pvt. Ltd

Fuel Prediction Application (FPA)

- Collected, collated and normalized existing consumption, performance and operational data of over **120k rows** from sensors and IoT devices
- Transformed raw data into **MySQL** with a custom-made **ETL** application to prepare unruly data for machine learning
- Developed machine learning models with **Python** and **scikit-learn**, which enabled the prediction of fuel usage with **95% accuracy**
- Successfully increased the **savings in fuel costs by 29%** using proprietary software, modelling tools and specialist information

Data Leveraging for Sales Conversion

- Designed, built and maintained data pipelines and systems for data ingestion from **APIs**
- Led a team of 4 interns to design an **end-to-end ML pipeline** for pre-processing (raw) client data, feature engineering, feature selection, model training and testing
- Updated data streamlining processes, that resulted in a **27% redundancy reduction**
- Built data models and maps within **AWS** environment to generate meaningful insights from client data, which resulted in enhancing sales efforts by **12% (revenue augmentation ~$300K, annually)**

Machine Learning Engineer   May 2018 – Aug 2018 | Noida, India

Tech Mahindra

- Built **recommendation solution** to provide personalized service to enhance user experience & improve ROI
- Implemented **Principal Component Analysis(PCA)** to identify the key business **KPIs**
- Used collaborative filtering to build a recommender system with Python, leading to augmentation of **150% repeat orders**

Education

MSc Data Science Feb 2021 – Feb 2022
University of Surrey Guildford, United Kingdom
Relevant Modules -Statistics, Machine Learning, Big Data, Natural Language Processing, Cloud Computing

MSc Applied Economics and Statistics (Online) Sep 2019 – Sep 2020
University of Strathclyde Glasgow, United Kingdom

B.Tech Computer Science and Electronics Engineering Aug 2015 – Aug 2019
Jaypee Institute of Information Technology Noida, Delhi (NCR), India

Data Science Projects

Stock Trading Chatbot Sep 2021 – present
Chatbot | Virtual Assistant | Conversational AI
- **Primary Goal**: Build a Stock Trading Assistant which analyses stocks and gives Buy/Sell recommendations
- **Proposed Solution**: The bot analyses stocks on 6 Technical Indicators and Market Sentiment Analysis
- **Result**: *Project under Development Phase*

Facial Expression Recognition using Deep Learning ⬀ Aug 2021
Classification Problem | Computer Vision
- **Primary Goal**: Build a web app to categorize facial expressions into seven different emotions
- **Solution**: Trained a CNN in Keras from scratch to recognize facial expressions in real-time
- **Result**: Achieved an accuracy of **73.85%**
- **Tools**: *Dataset - FER 2013 Data | CNN | Python | TensorFlow | OpenCV | Keras | Flask*

Movie Genre Prediction using Natural Language Processing ⬀ Jun 2021
Multi-Label Classification Problem | NLP
- **Primary Goal**: Build an interactive Web App to predict movie genres based on their plot descriptions
- **Solution**: Applied NLP techniques and Machine Learning Algorithms to classify movies as one or more genres
- **Result**: Achieved Precision Score of **88%** and Recall Score of **82%** using One vs Rest Classifier with Multinomial Naive Bayes using TF-IDF plot vectors (min_df=20) and Min-Max Scaling
- **Tools**: *Dataset - IMDB Movies | Python | SQL | Bag of Words | Standard Scaling | Logistic Regression | Flask*

Handwritten Digit Recognition ⬀ Apr 2021
Classification Problem
- **Primary Goal**: Build an efficient model to recognize handwritten digits
- **Solution**: Trained a CNN in Keras from scratch to recognize handwritten digits
- **Result**: Achieved an accuracy of 99.52%
- **Tools**: Dataset - MNIST Database | CNN | Python | TensorFlow | Keras

Customer Segmentation with Python Mar 2021
Unsupervised Learning | Clustering
- **Primary Goal**: Perform customer segmentation to help businesses understand their target audience
- **Solution**: Built a K-Means clustering algorithm for customer segmentation in Python
- **Result**: Successfully built a K-Means clustering model analyzed the behaviour of individuals in each cluster
- **Tools**: Dataset - Mall Customers Data | K-Means Clustering | Python | scikit-learn | NumPy | seaborn | matplotlib

Airlines Passenger Forecasting ⬀ May 2021
Regression Problem | Time Series
- **Primary Goal**: Develop a model to predict the number of international airline passengers for next 10 years
- **Solution**: Time Series Forecasting with **ARIMA** model in Python
- **Result**: Mean Absolute Error(MAE) of **10.05%** achieved.
- **Tools**: *Dataset - Air Passengers By Kaggle | Python | ARIMA | Matplotlib | Flask*

DataCamp Challenge : Promo Strategy for a Drinks Company ⬀ Nov 2021
Unsupervised Learning | Clustering
- **Primary Goal**: Target 10 regions that have similar buying habits to Saint Petersburg for a successful promotion
- **Proposed Solution**: Build a K-Means model to cluster the different regions based on the drinking habits
- **Result**: The next 10 regions for targeting wine promotions were successfully identified
- **Tools**: Dataset - Russian Alcohol Consumption | K-Means | Python | Standard Scaler