Vivekpandian Veerapandian

Sunnyvale, CA | vivekpandian08@gmail.com | (214) 802-4843
github.com/vivekpandian08 | website: vivekpandian08.github.io | linkedin.com/in/vivekpandian8

Data Scientist with **3-years of experience** in recommending data-driven business solutions by building Linear and Non-Linear models. Equipped with in-depth knowledge and practice of deploying scalable end-to-end machine learning models in the Cloud.

EDUCATION

M.S., Data Science and Business Analytics , The University of Texas at Dallas — GPA**:** 3.67 — **Dec 2020**

B.E., Electronics and Communication, College of Engineering Guindy**,** India — **May 2011**

TECHNICAL SKILLS

Programming : Python (Scikit-learn, Pandas, Numpy, TensorFlow, PyTorch, Keras, NLTK, PySpark), R, SQL, SAS

Data Visualization : Tableau, Power BI, Shiny, GGPlot2, Plotly, Matplotlib, Seaborn, Bokeh

Databases & Big Data : BigQuery, MySQL, PostgreSQL, Oracle, Graph, Redshift, MongoDB (NoSQL), Hadoop, Hive, Spark

Certifications : UDACITY Nanodegree (NLP, Data Engineering), Coursera(Deep Learning Specialization - **ANDREW NG**)

PROFESSIONAL EXPERIENCE

Senior Data Scientist, Ordermycake.in — **Jan 2015 – June 2018**

- Generated **18% increase in revenue to $5k** in 2017 by leveraging sentiment analysis, extracting topics and key phrases on customers feedback, and suggested a new payment method to solve delivery problems **[Python, NLP, AWS S3, EC2]**
- Identified **4500 potential churn customers** by developing ML models and retained 36% by offering them discounts **[R]**
- Revamped coupon mailing strategy for 3 customer segments by clustering using **K-means** and identifying the most engaging coupons leading to a 12% increase in headcount **[Unsupervised Learning]**
- Implemented marketing campaign using A/B testing to identify effective campaign, resulted in a **30% increase in sales**
- Recommended optimized price for products by web scraping and data mining to analyze competitor product prices, leading to a **$3K increase in yearly revenue**
- Led a 10-member cross-functional team to build an end to end B2C platform to expand the **operations from 2 to 8 cities**, and presented findings and insights to stakeholders **[Tableau]**

Software Engineer, Cluster Wireless Software Ltd — **Mar 2012 – Sept 2014**

- Designed **SQL queries** to extract information from IoT sensor and identified anomalies by K-means clustering **[MYSQL, C]**
- Procured project requirements, worked on sprint planning, root cause analysis, code reviews and collaborated with clients based out of different locations globally

INTERNSHIP EXPERIENCE

Data Science Intern, *SuperWorld*, United States — **Sept 2020 – Present**

- Predicted the likelihood of a purchase from clickstream data using Decision Tree and optimize the model by adding new features to improve the precision rate to 76% **[Python, Google Analytics]**
- Developed an interactive web app in R-shiny that does real-time **end to end sentiment analysis** of SuperWorld's tweets

PERSONAL PROJECTS (DATA SCIENCE)

Music Library Database Design, Data Modeling and Data Storage — SQL, Python, AWS, Redshift, S3, Spark

- Designed a star-schema based music database on Amazon Redshift, **built ETL pipeline** to extract raw data from S3 bucket into a staging area in Redshift and transformed it to build the database using data warehousing, SQL and Python
- Developed data lake for big-data analytics that reads data from S3, processes using Spark, and writes parquet files on S3

Document Classification using Deep Learning — Python, Natural Language Processing, GCP, PyTorch

- Leveraged web scrapping to scrape 30k news articles, extracted features by pre-training Word2Vec, BERT embeddings
- Implemented Convolutional neural network model to classify documents into 4 categories & achieved 92% accuracy

Vehicular Injury Predictive Model for New York City — Python, Statistics, Regression

- Performed hypothesis testing, deviance & multi-collinearity tests to estimate association of 19 features in road-collision, implemented Generalized Linear Model(GLM), and found 70% high risk for pedestrian & 92% risk for motorist

Credit Card Fraud Detection — Python, Supervised learning, Classification

- Performed quantitative analysis and feature selection**.** Experimented with algorithms like SVM, KNN, Naïve Bayes, Random Forest, and Neural Networks with **SMOTE** resampling using **PySpark** to predict fraud and achieved AUC of 0.82