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 extracting business insights from data points by building Linear and Non-Linear models. Equipped with in-depth knowledge and practice of deploying 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**

- *Coursework*: Statistics, Predictive Analytics, Econometrics, Time Series Forecasting, Machine Learning, Deep Learning
- *Project Mentor***,** BALC: Mentored 15 Grad Students in an Intra College ML project competition, and secured 3rd place

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

SKILLS

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

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

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

ML Concepts : Hypothesis Testing, A/B Testing, Regression, Classification, Clustering, NLP, Computer Vision

PROFESSIONAL EXPERIENCE

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

- Technologies: **Python, R**, MySQL, Tableau, AWS EC2, S3
- 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
- Designed 3 A/B tests to identify the most captivating marketing campaign, resulted in a **30% increase in sales**
- Revamped coupon mailing strategy for 3 customer segments by clustering using **K-means** and identifying the most engaging coupons leading to a **12% estimated increase in headcount**
- Identified **4500 potential churn customers** by developing ML models and mitigated 36% by offering them discounts
- 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**

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

- Technologies**: MySQL**, C
- Designed **SQL queries** to extract information from IoT sensor data and identified anomalies by K-means clustering
- Analyzed product pain points and collaborated with a multi-functional team to develop robust solutions to meet client requirements, increased project conversion to 30%

INTERNSHIP EXPERIENCE

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

- Technologies: R, Python, Google Analytics
- Built **Predictive model** using Decision Tree to find the likelihood of a purchase from clickstream data and optimize the model by adding new features to improve the precision rate to 76%

PERSONAL PROJECTS (DATA SCIENCE)

Stress Detection on Social Media Python (Natural Language Processing)

- Leveraged web scrapping to scrape 30k labeled Reddit posts and extracted features by pre-training Word2Vec, Doc2Vec and BERT embeddings with 190k unlabeled posts that capture semantic and syntactic similarity among words
- Trained **XGBoost** and **BERT** models to classify Stress posts on GCP, that achieved accuracy of 92.74% and recall of 94.58%

Traffic Sign Recognition for Autonomous Driving Python (Computer Vision)

- Pre-processed images and experimented with 5 different CNN architectures using Tensorflow to classify 43 traffic signs
- Deployed an interactive web APP that classifies traffic sign from user input using **Flask(REST API)** and **Kubernetes on GCP**

Credit Card Fraud Detection Python (Supervised Learning)

- Performed data analysis, and hypothesis testing**.** Built an **ML pipeline** using **python(PySpark)** to predict fraud transaction using SVM, KNN, Naïve Bayes, Random Forest, and Neural Networks with **SMOTE** resampling and achieved AUC of 0.82