

Vivekpandian Veerapandian

San Jose, CA (214) 802-4843
github.com/vivekpandian08

vivekpandian08.github.io
linkedin.com/in/vivekpandian8

vivekpandian08@gmail.com
medium.com/@vivekpandian08

Data Scientist with 3-years of experience in extracting customer and marketing insights from data points by building Linear and Non-Linear ML models. Proficient at framing ETL pipeline deploying end-to-end machine learning models in the Cloud.

TECHNICAL SKILLS

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

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

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

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

WORK EXPERIENCE

Data Science Intern | R Sept 2020 – Present

SuperWorld, United States – Augmented Reality Real Estate on Blockchain

- Developed an interactive web app in R-shiny that does **real-time end to end sentiment analysis** of SuperWorld's

Senior Data Scientist | Python, R, Tableau, AWS EC2, S3 Jan 2015 – June 2018

Ordermycake.in, India – E-Commerce B2C Platform to sell Bakery Products

- Generated **18% increase in revenue to \$5k** in 2017 by leveraging **NLP techniques** to analyze customers feedback
- 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 precision rate to 76%
- Designed 3 A/B tests to identify the most engaging marketing campaign, resulted in a **30% increase in sales**
- 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 reports to stakeholders using **Tableau** charts and dashboards
- Identified **4500 potential churn customers** by developing ML models and **mitigated 36%** by offering them discounts
- 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**
- Recommended optimized price for products by web scraping and data mining to analyze competitor product prices, leading to a **\$3K increase in yearly revenue**
- Initiated a new payment method to solve delivery problems that satisfied customers and **increased unit sales to 6.5%**

Software Engineer | MySQL Mar 2012 – Sept 2014

Cluster Wireless Software, India – Developed Software which facilitates M2M communications through IoT applications

- 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%

EDUCATION

The University of Texas at Dallas - **M.S., Business Analytics (Data Science Major)** 3.67 Jan 2019 – Dec 2020

- Project Mentor**, BALC: Mentored 15 Grad Students in an Intra College ML project competition, and **secured 3rd place**

College of Engineering Guindy, India - **B.E., Electronics and Communication** Aug 2007 – May 2011

PERSONAL PROJECTS (DATA SCIENCE)

Stress Detection on Social Media Python (Natural Language Processing) April 2020 – Aug 2020

- 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) Oct 2019 – Jan 2020

- 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) May 2019 – July 2019

- 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