KAVIN VEERAPANDIAN

Tampa, FL-33613 | +1(813)5604364 | [kavin.veerapandian@gmail.com](mailto:kavin.veerapandian@gmail.com) | [Linkedin](https://www.linkedin.com/in/kavin-veerapandian-b18537191/) | [GitHub](https://github.com/veerak2)

**WORK EXPERIENCE**

# Machine Learning Engineer – University of South Florida Jan 2021 – Dec 2021

* + Developed a machine learning system using Natural Language Processing to identify invalid assignment submissions which reduced the time needed for evaluation by 70%
  + Devised an implementation plan for a machine learning system including testing, deployment, and maintenance
  + Created a baseline model with a rule-based algorithm for binary classification with an accuracy of 55%
  + Identified that 10% of total data points were mislabeled through a heuristic method which affected model performance
  + Reduced the time taken to score an assignment by creating a front end with IPython widgets in a notebook
  + Developed a system to remove Personal Identifiable Information (PII) using models from Spacy with an accuracy of 95%
  + Created scripts in Python to extract, transform and load data into an Oracle database reducing 5 hours of weekly effort
  + Created custom trained word embeddings using word2vec from gensim which improved classification accuracy
  + Achieved reproducibility of experiments using versioning tools like Git, MLflow and Data Version Control
  + Implemented and evaluated multiple neural network architectures including RNNs and transformer-based models
  + Deployed the model using RESTful API with FastAPI and developed the front end for inference using Streamlit

# Teaching Assistant – University of South Florida Aug 2020 – Dec 2020

* + Held weekly office hours and mentored students on data modeling, cleaning, visualization, and interpretation
  + Lectured undergraduate classes on creating visualizations with Tableau for gaining insights from data

# Business Intelligence Engineer – ADP Feb 2016 – Feb 2019

* + Optimized data models & business rules resulting in increased loading time for reports built using an in-house tool.
  + Created dashboards for ad-hoc requests from internal and external clients for tracking KPI’s of application performance
  + Spearheaded full redesign of the employee onboarding module increasing engagement rate for module by 60%
  + Developed multiple SQL scripts in MSSQL server to extract information for a data driven product
  + Analyzed and identified key pain points in software development process reducing time taken for deployment
  + Designed workflows to automate the business processes for human resource modules using a .NET based in-house tool
  + Established data cleaning methods and increased data quality using validation techniques in Excel

**Education**

**MS in Business analytics and information systems Jan 2020 - Dec 2021**

University of South Florida, Tampa, FL

Course highlights: MLOps (Madewithml), Data Science Programming, Statistical Data Mining, Cloud Computing, Big Data Engineering

# Bachelor of engineering in computer science Aug 2011 - June 2015

KCG College of Technology, India

**SKILLS**

**Languages:** Python (Scikit-learn, TensorFlow, NumPy, Pandas, Spacy, nltk, etc.), R (Tidy verse, ggplot etc.)

**Tools:** VS Code, Jupyter, Databricks, MS Office Suite, Tableau, SQL, Spark, Git (Version Control), Docker

**ML models:** Regression (OLS, GLM, multi-level models, polynomial, time-series, Ridge, Lasso, Survival), Naïve Bayes, logistic regression, SVM, Gradient Boosting (XGBoost, Random Forest), Neural networks (LSTM, Convolutional Neural Networks, Transformers), Clustering (K-Means, DBSCAN, LDA)

**ML Tools:** MLflow (Experiment tracking), Optuna (Hyperparameter tuning), DVC (Data Versioning)

**Cloud:** AWS (EC2, Sagemaker, S3, RDS, EMR), Databricks (Community edition)

**Misc:** Feature Engineering, A/B testing, Data Visualization, Data Structures, Data Warehousing, ML system design

**ACADEMIC PROJECTS**

**Factors influencing homicides in the United States**

* + Gathered data from multiple government data resources i.e. annual reports available in Census, SAMHSA, FBI, SSTI, ATF and statefirearmlaws.org for the years between 2010 – 2019
  + Interpreted the results from a multilevel time series model to recommend actionable policies for reducing homicides

# Financial fraud detection using Spark structured streaming

* + Streamed synthetic transaction data from a Hadoop file system (HDFS) using Spark structured streaming in Databricks
  + Evaluated different classification algorithms using MLlib in Pyspark to identify fraudulent transactions on streaming data

# Produce classifier for self-checkout kiosks using computer vision

* + Implemented a convolutional neural network using functional API from Tensorflow in python
  + Used weights of produce as auxiliary inputs to classify similar looking produce which increased the accuracy by 8%