Manoj Venkatachalaiah
mv555@drexel.edu | (484) 442-0256 | LinkedIn | GitHub

Summary

- Graduate student at College of Computing and Informatics, Drexel University. Highly skilled in using statistical analysis, SSDT tools, data pre-processing, data mining and NLP algorithms, and working with database management to solve business problems. Experienced in networking; TCP/IP and socket programming. Proficient in Python, SQL and C++. Actively looking for full-time Data Scientist/ Analyst roles..

Education

M.S. in Computer Science | Drexel University | Philadelphia, PA (Graduation: *December 2021)*

- Minor: Machine Learning and Distributed Programming

M.S. in Data Science | Drexel University | Philadelphia, PA *September 2020*

- Minor: Machine Learning

B.S. in Computer Science | VTU | Bangalore, KA, India *June 2018*

- Minor: Database Management

Technical Skills

- Programming languages: Python, SQL (PL/SQL), C++, HTML, JavaScript, Scheme, Haskell, Matlab
- Machine Learning: Supervised Learning (Regression-Linear, Ridge, Lasso, Logistic | Decision Trees | SVM | Neural Networks | Random Forest | Gradient Boosting, XGBoost), Unsupervised Learning (K-Means | Hierarchical Clustering | Dimensionality Reduction | PCA | Cluster Analysis | Anomaly Detection), ANOVA, Market Basket Analysis, A/B and hypothesis testing, ARIMA, NLP: Text mining, Sentiment Analysis
- Data Analysis: Tableau, Power BI, Proficient in writing code (SQL and python) and using Big Data technologies (Hadoop, Spark ML, HDFS) to assess, clean, validate and analyze large datasets, SSDT: SSIS, SSRS, Expert knowledge of MS Excel; Proficiency in MS Access, MS Word

Academic Projects

Stock Prediction using NLP and Covid-19 data

- Used NLP to predict stock prices of Microsoft, Apple, Facebook and Amazon with a r2 score of 0.72, 10% better than average stock prediction projects.
- Features used were derived using over 10 years' worth of news data related to the 4 companies and world events (focused on the pandemic), previous stock data and Google trends data.
- Data acquisition, preprocessing and Predictive modelling was carried out using SSDT tools on Google Cloud Platform. EDA was carried out using python, Tableau and Power BI.
- News data was converted into sentiment scores using NLP. !4 features were used to predict the stock prices using LSTM, Linear Regression and Support Vector Regression algorithms.

Forecasting Confirmed Cases and Fatalities of Covid-19

- Forecasted confirmed cases and fatalities of Covid-19 using travel pattern and Google Trends data provided by John Hopkin's University.
- Used Linear Regression, Random Forest Regression and Decision Tree Regression algorithms to achieve a r2 score of 0.85 and RMSE of 2.36.
- Pipeline was created on Spark. Tableau, Pyspark, Spark SQL Spark ML were used in various stages of the project.

Analyzing the conditions contributing to Covid-19 Deaths

- Combined a CDC Covid-19 dataset and a public health dataset to study the effects of statewide features such poverty and literacy rates (among other features), medical conditions, age groups and mask mandate on Covid-19 fatalities.
- Used python, Tableau and Weka to explore patterns and see how each of the features lined up in contributing to Covid-19 fatalities.
- Performed feature selection and used the features to forecast Covid-19 fatalities with an r2 score of 0.78 and mean absolute error of 0.3. Algorithms used: Linear regression, Support vector regressor, Random Forest and decision tree regressor.

Work Experience

Teaching Assistant | Drexel University | Philadelphia, PA *January 2021 to Present*

- Courses: Data acquisition and pre-processing (DSCI 511) and Intro to Data Analytics (INFO 659)

Associate Event Manager | Hash Connect | Bangalore, KA, India *January 2016 to June 2018*

- Assisted in conducting recruiting events for top MNCs at Bagmane Tech Park.
- Responsibilities included advertising events, registering applicants, coordinating with company recruiters to identify suitable candidates.