Yeswanth Reddy Velapalem

315-278-2234 | yvelapal@syr.edu | www.linkedin.com/in/yeswanthr

EDUCATION

Master of Science in Applied Data Science, Syracuse University (3.87/4.0) **May 2021**

Relevant Coursework: Data Mining, Data Warehouse, Big Data Analytics, Text Mining, Artificial Neural Networks, Natural Language Processing, Financial Analytics.

Bachelor of Technology in Electronics & Communication Engineering **May 2018**

Panimalar Engineering College (3.33/4.0)

EXPERIENCE

Data Analyst, *i*Consult Collaborative, Syracuse University **August 2020 - Present**

- Collected, cleansed and validated the data of the Onondaga County Medical Society and maintained data accuracy to ensure accurate data availability for data analysis.
- Managed large datasets provided by the client using Index/Match and VLOOKUP and created dynamic reports and visualizations using Pivot tables and Charts on Excel.
- Assisted in creating a new website for the medical society which helped with automated data entry and report generation using Google Data Studio.

Data Science Intern, *Rubixe-AI* | Python, scikit-learn, Tableau **October 2018 – March 2019**

- Developed a predictive model classifier using python to classify the ticket priorities which reduced efforts of manual classification of the incoming tickets.
- Worked on ticket optimization, using predictive modeling techniques which led to 20% reduction in error rate and less ticketing.
- Optimized the model performance through parameter tuning, feature importance analysis and controlling the overfitting issue.
- Built a Tableau dashboard to forecast the incident volume in multiple fields, quarterly and annually.

ACADEMIC PROJECTS

Online News Popularity | Python: pandas, matplotlib | Apache Spark: PySpark, MLlib | Machine Learning | Statistics

- To classify whether an article is going to be a hit or not before publishing and predicting the share count of the article.
- Random Forest, Gradient Boosting and Logistic Regression is used to classify whether the article published by Mashable will be hit. Random forest was the best performing model with an AUC of 76% using PySpark.
- Random Forest Regressor model built to predict number of times an article will be shared with an acceptable RMSE.

Hospital Readmission of Diabetes Patients | Python: sklearn, pandas, matplotlib | Python Dash | Machine Learning

- Performed data wrangling and pre-processing to remove the data quality errors which owns to accurate predictions.
- Visualized the data to find the most relevant features which helped in the understanding how the features affected the re-admission of a patient.
- Built and fine-tuned a classification model by implementing Gradient Boosting and Artificial Neural Network algorithms to classify if a patient is readmitted or not.
- Analyzed model performance and fine-tuned Machine Learning models to maximize Accuracy & Recall.

Multivariate Time Series Weather Forecasting | Python: PyTorch, pandas, matplotlib | Deep Learning | Machine Learning

- Performed feature engineering on 14 weather-related variables with values recorded every hour for eight years.
- Integrated the predictive powers of ANNs with the data to forecast the temperature of each hour for next 24 hours.
- Developed a function to generate input sequences for past 4 days of data and output sequences with next day data.
- Improved the MSE for test data and time required for training using bidirectional LSTM networks over unidirectional RNNs.

TECHNICAL SKILLS

Programming Languages: R, SQL, Python, PySpark

Modules: Pandas, NumPy, Seaborn, matplotlib, scikit-learn, ggplot, dplyr, caret, SparkML, PyTorch, NLTK, Beautiful Soup

Concepts/Techniques: EDA, Data Visualization, Machine Learning, Deep Learning, ETL, Data Warehousing, Information Retrieval, Natural Language Processing, Web Scraping

Tools/Software: Word, Excel, Visio, Access, Power Point, Apache Spark, Jupyter Notebook, R Studio, MS SQL Server, Databricks, Tableau, PowerBI, AWS, Google Analytics

LEADERSHIP

Project Lead **August 2020 - Present**

- Leading a team of four members to tackle the data quality issues and curating organization data of the Onondaga County Medical Society.