pandey.v@northeastern.edu
LinkedIn: /in/viral-pandey

Viral Pandey

Available: Jan 2021 | (617) 955-0659

viralpandey.github.io
GitHub: /viralpandey

Education

Northeastern University Sep 2018 - Dec 2020
Master of Science in Data Science | ***GPA**: 3.9/4.0* Boston, MA
Graduate Teaching Assistant: Machine Learning, Data Management and Processing, Programming with Data
Relevant Coursework: Causal Inference, Deep Learning, Foundations of AI, Algorithms

Dhirubhai Ambani Institute of Information and Communication Technology Aug 2014 - May 2018
Bachelor of Technology in Information and Communication Technology India

Technical Knowledge

Languages: Python, R, SQL, Java, MATLAB, C, C++
Libraries: Pandas, Scikit-Learn, TensorFlow, Keras, Plotly, Matplotlib, Numpy, Pytorch, Pyro, OpenCV, H3, Geopandas
Statistical Methods: Time series forecasting, Hypothesis testing, Classification, Clustering, Regression Analysis, A/B test, NLP
Technologies: Airflow, Git, Jira, Hive, RStudio, Jupyter, Tableau, Power BI, Apache Superset

Work Experience

Tesla May - Aug 2020
Data Scientist Intern Palo Alto, CA

- Developed supervised regression models to predict congestion and determine the capacity expansion of Supercharger sites
- Built data pipelines to convert vector data of public roads into Uber's H3 hexagons. This helped me design and put Traffic Coverage and Road Coverage KPIs into production
- Identified vehicles that might be involved in potential misuse of the Supercharger network. Proposed false positive scenarios as well as solutions to mitigate such incidents

Tesla Aug - Dec 2019
Data Scientist Intern Palo Alto, CA

- Designed a time series forecasting model to estimate quarterly energy usage at sites. This informed the estimation of $ revenue from the entire Supercharger network for future quarters
- Quantified the population coverage of the world using geo-spatial data of population density per pixel of the world and isochrone coverage (areas within some minutes by driving) of sites

Northeastern University Jul 2019
Research Assistant Boston, MA

- Explored Procedure Learning to understand the constituting key actions of complex tasks from instructional video data
- Assembled a Fully Convolutional Sequential Network (FCSN) that produces a compact summary of the procedure steps and their ordering needed to perform a complex task, as well as localization of these steps in videos

Dhirubhai Ambani Institute of Information and Communication Technology Jan - Apr 2018
Data Science Research Intern India

- Outperformed other algorithms in forecasting Remaining Useful Life of a jet engine based on NASA's time series dataset by developing a Recurrent Convolutional Neural Network (RCNN) based predictive model

Projects

Named Entity Recognition (NER) and Relation Extraction (RE) from Patient's Medical Notes Sep - Nov 2020

- Highlighted entities like Drugs, Adverse effect, Dosage, Reason, etc and mapped the Drug entity with all other entities to create a structured data table out of unstructured notes
- Achieved 90% micro-F1 score for NER and RE using BioBERT and BiLSTM+CRF models
- Built a website and APIs to get model predictions using FastAPI and hosted them on Google Cloud Platform

Quora Insincere Question Classification Jan - Mar 2019

- Designed a supervised binary classifier to detect insincere content on the Quora website and compared performances of algorithms such as SVM, CNN and LSTM RNN
- Performed TF-IDF vectorization, Sentiment Analysis using Python NLTK framework for gauging overall sentiment

Bankruptcy Prediction Using Various Classifiers Jan - Mar 2019

- Fit Logistic Regression, Naive Bayes, LDA, QDA, SVM and feed forward Neural Networks on 5 years of Polish Companies Dataset containing 64 econometric ratios and bankruptcy labels
- Handled missing data using Mean value Imputation and class imbalance in training data using SMOTE
- Calculated Correlation matrix, performed Cross-validation for feature sub-setting and hyperparameter tuning