ABHISHEK PALLE

9792099153 | apalle1@tamu.edu | https://apalle1.github.io/ | apalle1

EDUCATION

Texas A&M University, College Station, TX — ***Dec 2020***
Master of Science, Industrial Engineering (Data Science Track) — *GPA: 3.6*
Courses - Machine Learning, Convex Optimization, Applied Analytics, Design of Experiments, Time Series Analysis, Deep Learning

Birla Institute of Technology and Science, Pilani, India — ***May 2016***
Bachelor of Technology, Mechanical Engineering — *GPA: 3.6*

TECHNICAL SKILLS

- **Tools:** Python, R, SQL, PySpark, AWS, Hadoop, Scala, Hive, SAP
- **Visualization:** Power BI, Tableau, bokeh, plotly, ggplot2
- **Deep Learning framework:** Pytorch, Keras, TensorFlow
- **ML package:** numpy, pandas, sklearn, scipy, statsmodels
- **ML:** Ensemble models, Object Detection models, Sequence models, Transformer models, A/B Testing, Multi-armed Bandit

WORK EXPERIENCE

Stanley Black & Decker, Data Scientist Intern — ***May 2019 - Dec 2019***
Business Intelligence - Reporting

- Developed analytics dashboard on Power BI to manage global assets, monitor field technicians, and track equipment utilization for decision-making purposes, thus successfully enhanced on-time delivery efficiencies by 16%
- Built a python bokeh interactive web application from scratch that helps visualize on-site deployment of equipment, technicians, and consignment movement over time for a given project, and hosted it on AWS EC2
- Generated refurbishment reports using SQL and visualized data in seaborn and plotly to get insights on variances in refurb costs across product lines, translating results into business actions

Customer Pricing Prediction

- Reduced RMSE by 38% while predicting consignment loss associated with customer order by identifying features from multiple sources
- Experimented with ensemble models like Random Forest, XGB, LGBM and interpreted predictions through error analysis, PDP, SHAP

Customer Transaction Prediction

- Analyzed customer segments and predicted future transactional behavior of customers with 78% accuracy using features like RFM
- Modeled lifetime value using BG/NBD model to predict the expected number of future purchases enhancing marketing efficiency

Time Series Forecasting & Anomaly Detection

- Built an ensemble time series forecasting model to improvise product demand forecast and reduced the forecast error, MAPE by 63%
- Detected anomalies accurately in real-time machine sensor data using DBSCAN, Local Outlier Factor, and Isolation Forest

Endless Robotics, Machine Learning Engineer — ***July 2017 - June 2018***
Object Detection for Wall-Painting Robot

- Tackled the problem of locating door co-ordinates in a user-supplied image by building an object detection model
- Built an image dataset, annotated the scraped images, and trained deep learning model (Faster-RCNN) using PyTorch
- Achieved a mAP@0.5 IOU of 0.78 while utilizing transfer learning and implementing data augmentation techniques

Production Modeling Corporation, Operations Analyst — ***July 2016 - June 2017***

- Performed discrete event simulation to analyze assembly line for a leading US automotive company; alleviated bottlenecks leading to an increase in throughput by 24 cars per hour
- Forecasted monthly demand of several product lines for India's largest steel manufacturer, performed simulation to recommend optimal equipment count and saved 300K USD in capital expenditure
- Designed an optimized supply chain network flow from distribution centers to customers using PuLP (python LP modeler)

RELEVANT PROJECTS

M5 Hierarchical Time Series Forecasting

- Engineered features and implemented LightGBM to predict daily sales of 30,490 products for Walmart for the next 28 days.
- Implemented recursive strategy to forecast multiple steps ahead, achieving a WRMSSE score of 0.67

Sentiment Span Extraction

- Fine-tuned different transformer-based pretrained models BERT, RoBERTa and ALBERT using PyTorch framework to extract word or phrase from the tweet that exemplifies the provided sentiment
- Experimented with preprocessing, model architecture, loss, training schedule, optimizer, postprocess to achieve Jaccard score of 0.72

Predicting Severity of Insurance Claims

- Predicted the loss incurred through an insurance claim based on 132 features using Ridge/Lasso regression, Random forests, and Xgboost
- Identified features using Factor Analysis of Mixed Data and obtained the least MAE of $1133 using Xgboost model

Wheat Head Detection

- Built a two-model ensemble of Faster-RCNN and EfficientDet to detect wheat heads, achieving an mAP@0.5-0.75IOU of 0.71
- Applied pseudo-labeling and image augmentation techniques such as CutMix, HSV, Flip to further improve the metric by 6%

Multi-Label Topic Classification of Tweet

- Extracted tweets from twitter using Tweepy API for different combinations of topics
- Implemented ensemble of classifier chains to extract all the topics a user is talking about in a tweet with a precision score of 97.2%

Credit Card Fraud Detection

- Analyzed an imbalanced dataset and improved the recall by 42% using over-sampling technique (SMOTE)
- Performed hyper-parameter tuning on Logistic Regression, Ensemble methods and SVM to identify best solution for the classification task