Prathmesh Savale

https://prathsoo7.github.io/ • https://github.com/prathsoo7 • prathmesh.savale@gmail.com • +91 8087744932

Experience

Kiewit Corporation — Bangalore, India
Data Science Consultant — *Oct '18 – present*

- **Predictive maintenance of Caterpillar haul trucks at coal mines**
  - Built a classification framework using LSTM networks to predict breakdown of Caterpillar haul trucks at Kiewit's coal mining facility. Accurate prediction helps save additional cost of repairing an unplanned breakdown and increases throughput of the mining facility.
  - Improved the AUROC by 20% using rolling, tumbling and hopping aggregates on the truck's sensor data.
  - Used SHAP values to identify features that contribute to the maximum number of breakdowns.
  - Deployed the solution on premise using tensorflow/serving and flask API.
- **Text classification for invoice verification of construction and engineering procurement**
  - Created a text classification framework to tag commodity procurement invoices with SAP Ariba taxonomy. Correct identification helps verify expenditure and identify fraudulent supplier claims.
  - Used iterations of GloVe, Word2vec and custom layer embeddings to generate word vectors for classification in a deep neural network that achieved an AUROC of 92%.
- **Route optimization of construction vehicles across rough terrain**
  - Created an application using elevation raster, Haversine distance and routing algorithms to find the best path for traversal of construction vehicles across rough terrain for ease of operation and fuel efficiency.

Mu Sigma Inc. — Bangalore, India
Decision Scientist — *Sep '15 – Oct '18*

- **Building sales forecasting framework | Client - UK's largest retailer**
  - Built a forecasting framework using time series based ensemble model to produce forecasts at different levels of the buying hierarchy. It is used by commercial teams for budgeting and inventory management.
  - Included adjustment for external regressors like holidays and used forecast scaling for store closures which helped improve the overall company level forecast accuracy by 5.6%.
  - Used dynamic time warping to cluster similar time series shapes which helped decrease runtime of grid search for hyperparameter tuning. Parallelized model building, scoring, and forecasting for ~2500 stores and ~3600 products using PySpark. Used test-driven development to ensure error free codebase in a CI/CD pipeline.
- **Reducing device failure rates | Client - Fortune 3 technology company**
  - Created a boosted trees ensemble to predict electronic device failures leading to a 3% reduction (9% to 6%) in failure rate which translates to a cost reduction of ~1.8 million USD annually in inventory management.
  - Implemented cascading classifiers to decrease collateral damage while predicting device failures.
  - Completely automated and deployed the analytical solution using Jenkins saving ~40 man-hours each week.

Persistent Systems — Pune, India
Engineering Intern — *Jun '14 – May '15*

- **Developing CUDA based image processing application | Internship**
  - Developed a CUDA C based application to execute a computationally expensive content-aware image resizing algorithm called seam carving on GPU. This helped achieve 7.5X acceleration in execution time over traditional CPU execution due to the high degree of parallelism of Nvidia proprietary CUDA based GPUs.

Education

University of Pune — Pune, India
Bachelor degree in Computer Engineering | First class with Distinction — *2011 – 2015*

Skills

Programming: Python and PySpark, C and C++, R, SQL and HiveQL, Bash
Computational programming: octave, numpy, pandas, seaborn, scikit-learn, statsmodel, keras, nltk, gensim
Statistical analysis: Regression, Bagging, Boosting, Ensemble, Hypothesis testing
Tools: Jupyter, Teradata, Spark and Hadoop, Git and Github, LaTeX, Jenkins, Docker, Jira

Certifications

Machine Learning: Audited Coursera MOOC by Andrew Ng. [Certificate] [Code]
Data Science Math Skills: Audited Coursera MOOC by Duke University. [Certificate]
Decision Scientist: Audited certification course by Mu Sigma. [Certificate]
Machine Learning A-Z: Hands-On Python & R In Data Science: Udemy MOOC. [Certificate]

Personal Projects

Tech support call logger: Using LIUM for speaker diarization and google speech API for speech transcription. [Code]
Super Mario bot: Using Deep Q-learning to train agent in the NES classic super mario bros environment. [Code]
dabl: Open source contribution to the data analysis baseline library based on scikit-learn for autoML. [Code]