Onkar Deshpande

San Mateo, California
onkardeshpande.github.io/ OnkarDeshpande onkar-deshpande-1207
onkar.deshpande@stonybrook.edu
+1 (631) 820 5236

EDUCATION

Stony Brook University, NY
Master of Science, Applied Mathematics and Statistics August 2017 - May 2019
Coursework: Probability, Data Analysis, Regression, Categorical Data Analysis, Mathematical Statistics, Big Data, Survival Data Analysis, Design of Experiments, Linear Algebra

Indian Institute of Technology, Kharagpur, India
B.Tech-M.Tech (Honours) Industrial Engineering July 2010 - May 2015

TECHNICAL SKILLS

- Programming Languages - Python (Pandas, Numpy, Scikit-Learn, SciPy, Keras, PyTorch, Matplotlib), R, MATLAB
- Machine Learning - Linear / logistic Regression, Regularization, K-nearest neighbors, Decision Trees, Ensemble Methods (LightGBM, Random Forests), Support Vector Machines, Clustering (Hierarchial, K-means), Principle Component Analysis, Neural Networks, NLP, Predictive Models; Statistical Methods: ANOVA, Hypothesis Testing, A/B testing
- Tools and Databases - Git, HTML, Google Cloud Platform, MongoDB, Kafka, Spark, MySQL, Teradata

EXPERIENCE

Express Scripts Inc, St Louis, USA — *Data Scientist - Intern* May 2018 - August 2018

- Built 3 therapy level linear predictive models (AUC >0.8) for member engagement; part of clinical technology product
- Engineered new features to accommodate for historical alterations in member medical claims data (1.4B data points)
- Projected revenue of $ 20M in the next two years; the models will go in production in January 2019

Affine Analytics, Bangalore, India — *Data Scientist (client,* ***Sears Holdings****)* November 2016 - July 2017

- Implemented gradient boosting method (LightGBM) to build a logistic classifier (Dataset size: 1 Million) to output purchase propensity of the member to visit store next month for a business unit
- Engineered features by building embeddings to encrypt buying pattern of a member, improved model accuracy by 4%
- Built a gradient boosting model to capture the member engagement over mailers sent for SYW, Sears & K-Mart

Robert Bosch, Bangalore, India — *Senior Analytics Engineer* July 2015 - November 2016

- Developed two stage SVM model (0.92 precision, 0.96 recall) to identify anaemia severity using data from non-invasive sensors; Used Agglomerative clustering to tackle class imbalance ; John Hopkins University Research
- Identified Anomalies in vehicle test data files using Kullback-Leibler divergence; Implemented the solution in Apache Spark; Client noted a 3.5-4x speed improvement against their current in-place solution
- Engineered a regression model to classify the dispensed beverage through signal taken from the vibrations of coffee vending machine and predicted raw material requirements; Used Independent Component Analysis (ICA) technique to remove noise from the data
- Worked on a text mining platform to build modules to clean text (lemmatization, punctuation & numbers removal) and extract features (Sentence body detection, POS-tagging, key words, Chunk parser); Used nltk & gensim libraries in python (English language); The tool was hosted on internal cluster that be accessed using API call
- Built a tool for internal use in R to accommodate for forecasting techniques like exponential smoothing, ARIMA and Holt-Winters along with data preprocessing modules

PROJECTS

Sequence Classification

- Trained and compared a RNN (bidirectional LSTM) vs a 1-D CNN model on TREC dataset to classify given question into six different categories; The model achieved accuracy of 81%

deeplearning.ai Completed Deep Learning specialization on Coursera November 2018 - February 2019
Neural Networks, Deep NN, Convolutional NN, Sequence Models

Kaggle: Competition Expert *Top 3% global ranking*

- Elo Merchant Category: Top 3% (4000 Teams); Predict customer loyalty using customer transaction history
- Santander Customer Transaction: Top 6% (9000 teams); Binary classification - customer will make transaction in future

Ideas implemented: Model Stacking / Blending, LightGBM, Random Forests, Linear Regression with ElasticNet, Stratified Cross Validation, Feature engineering and aggregations, Data Augmentations.