Onkar Deshpande

700 Health Sciences Dr, Chapin I 1135, Stony Brook, NY 11790 | onkar.deshpande@stonybrook.edu
onkardeshpande.github.io/ | Github | LinkedIN | +1 (631) 820 5236

OBJECTIVE

Looking for intern opportunities in Data Science and Machine Learning, where I can apply my 2+ years experience in predictive analytics as well as leverage my background in statistics.

EDUCATION

Master of Science, Applied Mathematics and Statistics 2017 - 2019 (Expected)
Stony Brook University, NY
Coursework: Probability, Data Analysis, Regression, Categorical Data Analysis, Mathematical Statistics

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

TECHNICAL SKILLS

- Programming Languages - Python (pandas, numpy scikit-learn, matplotlib), R, MATLAB, SQL
- Machine Learning - Decision Trees, Ensemble Methods (LightGBM, RF), Support Vector Machines, Linear/logistic Regression, Clustering, Dimensionality reduction, Neural Networks, NLP
- Tools and Databases - Git, HTML, Google Cloud Platform, MongoDB

EXPERIENCE

Data Scientist November 2016 - July 2017
Affine Analytics, Bangalore, India

- Worked for client, **Sears Holdings**
  - 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 new features by building embeddings thereby encrypting the buying pattern of a member, improving model accuracy by 4%
  - Built a gradient boosting model to capture the member engagement over mailers sent for Shop-Your-Way, Sears and K-Mart separately

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

- Implementing supervised and unsupervised Machine Learning algorithms to solve different business problems
  - Developed a two stage SVM model (0.92 precision, 0.96 recall) to identify anaemia severity by using data from non-invasive sensors; Used agglomerative clustering to tackle class imbalance in the data; Part of research project at John Hopkins University
  - Identified Anomalies in vehicle test data files using Kullback-Leibler divergence; Implemented the solution in Apache Spark
  - 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
- Building tools useful for the department as a part of Centre of Excellence team
  - 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); Hosted the tool on the cluster which could 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

Hypothesis testing Data Analysis

- Analyzed credit card data for different hypothesis testing; Used ANOVA, Multivariate regression to identify factors for credit limit, default rate and analyze buying patterns over six months period.

Movie Recommender Fundamentals of Computing

- Implemented machine learning models from scratch on movie-lens data set. Compared performance of three different models: Collaborative filtering, K-means, hybrid model.

Dissertations, IIT Kharagpur Guide Prof. P. K. Ray

- Used Holt-Winters model to improve throughput of the inventory management system by 3%; Optimized the inventory replenishment process of spares and consumables at TATA Bearings, Kharagpur
- Modeled product scheduling with machine breakdown and normally distributed processing time; Employed evolutionary search algorithm, to arrive at schedules with reduced variability of 8% compared against genetic algorithm