Onkar Deshpande

700 Health Sciences Dr, Chapin I 1135,
Stony Brook, NY 11790

onkar.deshpande@stonybrook.edu
+1 (631) 820 5236

INTERESTS

Data Mining, Machine Learning, Natural Language Processing, Predictive Analytics

EDUCATION

Master of Science, Applied Mathematics and Statistics 2017 - 2019 (Expected)
Stony Brook University, NY
Current Coursework: Introduction to probability, Data Analysis, Linear Algebra and Advanced Calculus

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

COMPUTING SKILLS

Python, R, MATLAB, Git, MySQL, PostgreSQL, MongoDB, NumPy, Pandas, Scikit-Learn, Matplotlib, LaTeX

EXPERIENCE

Data Scientist November 2016 - July 2017
Affine Analytics, Bangalore, India
Worked for client, **Sears Holdings**

- Identifying members visiting the retail stores monthly
  - Implemented gradient boosting method (LightGBM) to build a logistic classifier that outputs 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%
  - Worked on building 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 the vehicle test data 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
  - Simulated thousands of vehicles under a python environment to generate data input to the big data architecture via Kafka. This model was used to organize an internal Hackathon on Vehicle Telematics
  - 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

Data analytics intern
IPSOS Research Pvt.Ltd., Bangalore, India May 2014 - June 2014

- Created & developed a Market Mixed Modelling VBA tool to estimate sales over marketing spend across multiple levels of geography, Achieved precise results with an error of less than 3% after validating output against that of R & SAS
- Worked on open source R Platform to obtain and fit sales models for Maybellines products - Big Eyes, Mega Plush

ACADEMIC PROJECTS

M.TECH. Dissertation, IIT Kharagpur Guide Prof. P. K. Ray

- Used Holt-Winters model to improve throughput of the inventory management system; Optimized the inventory replenishment process of spares and consumables at TATA Bearings, Kharagpur

B.TECH. Dissertation, IIT Kharagpur Guide Prof. P. K. Ray

- 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

Work Authorization: Eligible to work in USA under CPT.