SHARAN SUBRAMANIYAN

| 2752 N Seminary Ave Apt 2F, Chicago, IL 60614 | 443-515-8339 | ssbrmny2@illinois.edu |

Graduate in Statistics (Analytics) with a passion to apply data analysis, statistics and machine learning techniques to tough business problems to help unlock growth and profitability.

EDUCATION:

University of Illinois at Urbana Champaign *Jan 2016 – Dec 2017*
Masters in Statistics (Analytics) GPA – 3.45 / 4.00
M.S. Ramaiah Institute of Technology, Bangalore *June 2011*
Bachelor in Engineering – Chemical Engineering GPA – 9.16 / 10.00

WORK EXPERIENCE:

Data Science Intern, John Deere *May 2017 – Dec 2017*

- Used natural language processing methods to automate call center activity by mapping solutions to case text
- Data used for this was human transcribed free form text data. Data was prepared for analysis using data cleaning techniques like stop word removal, lemmatization, tokenization, word segmentation, xml and html parsing.
- Used various vectorization techniques like word2vec, doc2vec, tfidf
- Similarity search using cosine similarity distance measure
- Worked on classification using Random forest, SVM
- Clustering using k means and density based clustering methods like DBSCAN, HDBSCAN, OPTICS.
- Dealt with missing labels for image classification and used Amazon mechanical turk to set up data labelling tasks
- Used python packages like spacy, genism, nltk, pandas, numpy, sklearn, bs4, xml for the above project.
- Worked on a classification problem to classify feedback text data into pre existing buckets using random forest and neural network approaches.

Data Analyst (Grad Hourly), Facilities & Services, University of Illinois at Urbana Champaign *Feb 2016 – Dec 2016*

- Actively consulted with management and various teams in the organization to gather user and data requirements to create the following applications:
- Scheduling application to automate and boost efficiency of building work order management.
- Decision support application to improve timing of building visits, to identify faulty meters, measure a building visit impact in terms of ROI along with an automated dash board using VBA, userforms and excel chart object programming that updates itself on a daily basis.
- Tools used: Advanced excel (pivot tables, data validation, solver, user forms, macros), wrote custom programs using Excel VBA with integrating database querying from within excel using connection strings to connect to MS SQL Server and MS Access databases. Wrote complex SQL queries involving joins, merging and sub- querying. Also used power pivot with DAX formulae to achieve certain outcomes.

Systems Trader, Percentage Play, Bangalore *Sept 2014 – Dec 2015*

- Created company MIS, backtested trading systems with SAR system resulting in over 30% profit in an account.

PROJECTS:

Data mining WSDM 2017 competition (Triple Scoring)

- Used word2vec to create word embeddings from wiki text corpus. Data cleaning, data visualization. Used support vector regression, Gradient boosting, ensemble to train models. Tried unsupervised learning with cosine similarity
- Predicted relevance score of given triple (person – profession or person – nationality) on test data.
- Achieved 79.72% accuracy and placed 3rd (out of 21 teams) in the competition (team *radicchio*)

Big data: Airline data analysis

- Data subsetting and merging using Hive in Hadoop followed by data visualizations such as tree maps, heat maps.
- Prepared data in pandas using chunk wise data import (using generators) and used stochastic gradient descent classifier to learn a classifier on large data. Obtained an accuracy of 76% when predicting arrival delays.

Machine learning applied to trading

- Feature engineering to create predictors. Used regression (elastic net, KNN) and classification (QDA & LDA, SVM, KNN, Random Forest) algorithms to train models, and back tested their performance using past market data.
- Predicted daily return and performed better than the S & P 500 index for the given time period.

Home Price Genie

- Created a data product integrating R (backend engine), Excel (dashboard) and VBA (macros with shell interfacing)
- Performed diagnostics on the house price data (outlier analysis, leverage point analysis, Box Cox transformation) and used full linear model, variable selection model (AIC BIC) and random forest to predict house prices.

Implementation of Algorithms from scratch using python:

- Apriori algorithm for mining frequent itemsets
- Decision Tree & Random Forest classification algorithms

PROGRAMMING LANGUAGES:

- Python, SQL, R, Excel VBA
- Exposure to SAS, Java, Hadoop (Pig, Hive)

COURSE WORK:

- Statistics and Probability II
- Basics of Statistical Learning
- Statistical Data Management
- Introduction to Data Mining
- Applied Regression and Design
- Statistical Learning
- Data Mining Principles
- Data Science Foundations
- Mathematical Statistics I
- STAT Internship
- Methods of Applied Statistics