Satvik Shetty

✉: smshetty@ncsu.edu | ✆: +1 919-771-6547 | ⌖: Raleigh, NC- 27606 | linkedin.com/in/satvikshetty | github.com/satvikshetty04

ACADEMICS

Master of Computer Science	North Carolina State University	May 2018	GPA: 4.0
Bachelor of Engineering in Computer Engineering	University of Mumbai, India	June 2014	First Class w/ Distinction

Key Courses: Database Management System, Automated Learning & Data Analysis (ALDA), Design & Analysis of Algorithms, Data Guided Business Intelligence, Advanced Data Structures, Artificial Intelligence, Foundations of Data Science, Introduction to Big Data, Data Intensive Computing, Inferential & Descriptive Statistics, Data Driven Decision Making

TECHNICAL SKILLS

Languages: Python, R, Java, SAS, C++, C, HTML, CSS, JavaScript
Databases: PostgreSQL, MongoDB, Oracle 12c, MS SQL, MySQL
Data Visualization / Reporting: Tableau, Pentaho, Microsoft Excel, Power BI
Tools / IDE: RStudio, PyCharm, Eclipse, Android Studio, SAS Studio, Jupyter, Git
Frameworks: Apache Spark, Apache Kafka, Apache Hadoop

PROJECTS

Machine Learning Challenge: Predicting AD clicks *[Tableau, Python]*
- Predicted the probability of an AD being clicked using Extreme Gradient Boosting (XGBoost) and CatBoost Classifiers
- Visualized data using Tableau to understand feature significance and performed the ensembling of models in Python

Multi-label Text Classification *[R, Python]*
- Performed semi-supervised classification on The Guardian news data-set into seen and unseen topics
- Employed various Natural Language Processing (NLP) techniques on text data for feature generation
- Analysed the performance of various ML classification algorithms using R's mllib package and Python's scikit package

Attributed Graph Community Detection: Market Segmentation *[R, Python]*
- Implemented Structure-Attribute Clustering algorithm for attributed graphs to identify the relevant market segments
- Evaluated the obtained segments using influence propagation in R

Prediction using Bayesian Regression *[Python]*
- Predicted the price variations of Bitcoin in Python using scikit, numpy, pandas and statsmodels packages

Twitter Sentiment Analysis *[Zookeeper, Apache Kafka, Apache Spark Streaming, PySpark]*
- Performed sentiment analysis on real-time tweets using Python, Apache Spark Streaming, and Apache Kafka

Recommender System using Collaborative Filtering *[Jupyter, Apache Spark, PySpark]*
- Recommended new artists to a user based on listening history on an audioscrobbler dataset using Apache Spark and Python

Database Application & SimpleDB *[Java, Oracle 12c]*
- Built a simple Swing (Java) application with Oracle 12c (backend) to track/manage patient health status
- Implemented FIFO buffer replacement policy and Undo-Redo recovery algorithm in the existing SimpleDB server

Predictive Analysis: Player Transfers in Soccer *[R, Python, SQLite]*
- Built a player transfer prediction model using K-means and DBSCAN clustering in Python/R as part of the ALDA course
- Predicted the outgoing (weakest player) and incoming players (replacements) for a team through statistical analysis in R

WORK EXPERIENCE 2 YEARS

EdgeVerve Systems Limited, Pune, India **Aug '15 – Jul '16**
Product Integration Engineer / Business Intelligence Analyst

- Consolidated, verified, and analyzed data received from distributors for major FMCGs across various geographies
- Generated sales, forecast and inventory reports to help visualize, strategize, and understand the data via the tool Pentaho
- Optimized the target data extraction query in PostgreSQL by 90% - brought down the run time from 3 hours to 10 mins
- Part of Android development team to create an application for automatic offline data retrieval component

Infosys Limited, Pune, India **Jun '14 – Jul '15**
Systems Engineer / Data Analyst

- Wrote stored procedures in MS SQL Server for data rectification, validation, and extraction
- Performed data analysis using Excel and MS SQL to ensure correctness and uniformity despite differences in ERP systems
- Developed and maintained a Java based online retail store application called 'JCart', as part of the training in core Java

ACHIEVEMENTS

- Ranked 9th among 5567 in the HackerEarth Machine Learning Challenge in predicting the probability of AD clicks
- Among the top 5% in the *Growing Instability* Challenge *(*Multi-Label Text Classification*)* on datasciencechallenge.org