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: 3.96
Bachelor of Engineering in Computer Engineering	University of Mumbai, India	June 2014	First Class w/ Distinction

Key Courses: Automated Learning & Data Analysis, 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, Object Oriented Design & Development

TECHNICAL SKILLS

Languages: Python, R, Java, SAS, C++, C, HTML, CSS, JavaScript
Databases: PostgreSQL, MongoDB, Oracle 12c, MySQL, MS SQL
Tools / IDE: PyCharm, Eclipse, Android Studio, RStudio, SAS Studio, Jupyter, Git, Tableau, PowerBI
Frameworks / Platforms: Django, Bootstrap, Apache Spark, Apache Hadoop, AWS

PROJECTS

Scalable Search Engine *[Python, AWS]*
- Built an almost real-time scalable search engine pipeline capable of handling high volume of text/news articles and high velocity of queries. AWS Kinesis, Lambda, S3 and ElasticSearch (indexing) were used for the same
- Integrated a gradient boosted model to automatically classify documents into categories to include contextual information

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

Health Insurance Selector *[Python, Django, SQLite]*
- Developed a web application in Django to return the best health insurance plans based on limited user inputs
- Designed a modified k-NN approach for plan matching and performed sentiment analysis on web scraped reviews

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

Truth Maintenance System *[Python]*
- Built a rule based truth maintenance system that would update its knowledge base based on a sequence of operations

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

Commutator *[Android, Java, SQLite]*
- Created an Android application based on the concept of dynamic ridesharing by allowing users to arrange an ad-hoc ride
- Developed the user interface on Eclipse and managed database connectivity through SQLite and Parse for cloud storage

WORK EXPERIENCE 2 YEARS

NC State - CIPM, Raleigh, NC **Oct '17 – Current**
Programming Intern
- Developed a web application using Django to assist the USDA in performing the spatiotemporal risk analysis of Fruit Flies

EdgeVerve Systems Limited, Pune, India **Aug '15 – Jul '16**
Product Integration Engineer / Business Intelligence Analyst
- Analyzed & processed data and generated reports to help visualize, strategize, and understand the data via 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
- Performed data analysis & wrote stored procedures 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