Simaant Patil

sjpatil@syr.edu +1 315 395 9581 Syracuse, NY 13210
https://www.linkedin.com/in/simaantpatil https://github.com/simaant

EDUCATION

Master's in Information Management, Syracuse University **GPA**: 3.78 **May 2020**
Certificate of Advanced Study in Data Science
Courses - Big Data, Database Management, Data Science, Natural Language Processing, Business Analytics, Project Management
Bachelor of Technology in Electronics, University of Mumbai **GPA**: 3.69 **June 2018**
Relevant Coursework - Data Analysis, Statistics, Neural Networks and Fuzzy Logic, Data Structures, Cloud Computing

SKILLS

Programming Languages - R, Python, SQL, Spark, Scala, HTML5, CSS, JavaScript, Node.js
Databases - PostgreSQL, Apache Cassandra
Libraries - NumPy, Pandas, Matplotlib, Plotly, Keras, PySpark, SciPy, TensorFlow, Scikit-learn, PyTorch, ggplot2, dplyr, tidyverse
Machine Learning - Regression, Clustering, SVM, PCA, Decision Trees, Random Forest, XGBoost, Neural Networks
Tools - Tableau, Git, MS Excel (DAX, Pivot Tables, Power Query), SAS, Jupyter Notebook, AWS, Google Analytics, JIRA, Power BI

EXPERIENCE

Business Data Analyst Intern, iConsult Collaborative / City of Syracuse, Syracuse **Feb 2020 - May 2020**
- Performed data extraction of complaints and violations of rules in the Syracuse area by employing Python and R scripts
- Developed SQL scripts in SQL Server and performed data cleaning and analysis using DAX functions in Power BI
- Incorporated Power BI to define relationships and create visualization reports to regulate the cause of complaints and violations

Graduate Research Assistant, Martin J. Whitman School of Management, Syracuse University **Feb 2020 - May 2020**
- Compiled data of 1M domain names in R using to determine medical websites and to understand user behavior
- Utilized a glossary of medical terms to scrape the websites and textual content present in the websites using a list of HTML tags
- Applied topic modeling and Google Analytics to determine patterns of user behavior and providing suggestions to them

Data Analyst, iConsult Collaborative, Syracuse University **Feb 2019 - May 2020**
- Collaborated with cross-functional teams to explore the generated patient data for finding disease patterns in various counties
- Incorporated SSIS packages and SQL for ETL and improved the overall efficiency of the process by 20%
- Designed interactive Tableau dashboards and generated maps of the results depicting the diseases prevalent in each county

Graduate Machine Learning Researcher, NEXIS Student Technology Lab, Syracuse University **Aug 2019 - May 2020**
- Acquired data of 8M Windows machines in Python for finding the probability of them being affected with malware
- Applied Apache Spark for performing feature engineering and extracting most relevant attributes like OS type, RAM, antivirus
- Created Tableau visualizations for detecting trends and implemented models such as SVM, logistic regression with 72% accuracy

PROJECTS

Diabetes Disease Data Classification - Big Data Analytics
- Analyze data from 1990-2008 of 100K values with 50 features to identify a patient's chance of being readmitted due to diabetes
- Performed data cleaning for handling null values and feature engineering for developing interpretable data from raw values
- Created a PySpark pipeline with PCA for determining the components from the features such as age, disease type, etc.
- Utilized the pipeline to develop classification models such as Random Forest with 70% precision in the results

Quora Insincere Questions Classification - Natural Language Processing
- Leveraged data exploration techniques using Python to identify toxic and divisive questions posted on Quora by its users
- Carried out data cleaning and data wrangling using NLTK as per NLP rules for extracting important features from textual data
- Implemented predictive modeling using logistic regression, SVM and deep learning models such as LSTM and CNN
- Generated visuals of the data trends using Matplotlib and predicted the output with a model accuracy of CNN with 85%

Prescriptive Analysis for Airline Companies - Data Science & Analytics
- Analyze data of 130,000 customers in R for understanding declining customer count for the airline companies in the U.S
- Performed data manipulation and validation using R and utilized Tableau for identifying KPIs like age, gender, etc.
- Predicted the features affecting the profits by exploiting techniques such as SVM and apriori algorithm with 75% accuracy
- Generated a report from Tableau dashboards and results from machine learning models to improve market presence

Data Warehouse for Fudge Corporation - Data Warehouse and Business Intelligence
- Utilized data of two companies, an online retailer and a DVD rental company to develop the company's data warehouse using ETL
- Modeled the staging area for extracting data from the source using SSIS packages and verified the structure in MS SQL Server
- Developed SSAS cube to create data hierarchies and generate KPI's for improving sales, customer reviews and delivery time
- Created Power BI dashboards using data from cube and data warehouse to develop business insights to improve overall profits

LEADERSHIP

Data Science Program Manager, NEXIS Technology Lab - Managed 10 data science projects for the research lab
Finance Chair, iSGO, Syracuse University - Organized networking events and advising sessions for graduate students
Graduate Mentor - Provided guidance to graduate students with respect to curriculum and professional development