

Simaant Patil

sjpatil@syr.edu

+13153959581

Dayton, OH, 45402

<https://www.linkedin.com/in/simaantpatil>

<https://www.simaantpatil.com>

<https://github.com/simaant>

EDUCATION

Master's in Information Management and Data Science, Syracuse University **GPA: 3.78** **May 2020**

Courses - Big Data, Database Management, Data Science, Enterprise Risk Management, 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

Cloud Technologies: AWS (Redshift, S3, EMR)

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, Neural Networks

Tools - Tableau, Git, MS Excel, SAS, Jupyter Notebook, Google Analytics, Power BI, SharePoint

EXPERIENCE

Data Scientist, Ascend Innovations, Dayton, OH **Jan 2021 - Present**

- Acquired assessment and demographic data of patients from a SQL database and performed feature engineering in Python
- Utilized features like age, gender, and answers from assessments from the data for exploratory analysis using Plotly
- Employed tree-based machine learning models for predicting short-stay patients, length of stay, and cost incurred in hospice care
- Developing a web application using Flask API for pulling results for short-stay patients having an F1 score of 0.82
- Testing deep learning methods for the length of stay and cost variables and Power BI to develop visuals using Flask API

Data Science 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 by 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

Mobile Transactions Fraud Detection - Big Data Analytics

- Utilized synthetic mobile transactions data of 6M instances from Kaggle to investigate fraudulent activities between accounts
- Cleaned the dataset using Python to remove nulls, standardizing and normalizing values in 8 out of the 11 features
- Used stratified K fold sampling method to deal with the unbalanced dataset to develop accurate training and evaluation sets
- Implemented Random Forest with Grid Search, SVM, and assessed the model performance using precision, recall, and F1 score
- Generated a classification report and incorporated the best Random Forest Model with a F-1 score of 0.86

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

- Analyzed 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

Finance Chair, iSGO, Syracuse University - Organized networking events and advising sessions for graduate students