Sohaib Syed

Lincolnwood, IL 60712 • sohaibsyed21.github.io/ProfessionalWebsite/ • sohaibsyed6@gmail.com • linkedin.com/in/sohaibsyed6/

EDUCATION

Illinois Institute of Technology, Chicago, IL   May 2024
- Bachelor of Science, Computer Science, **3.92**/4.0
- Master of Applied Science, Data Science, **3.90**/4.0

Relevant Courses: Machine Learning, Applied Statistics, Data Mining, Data Prep & Analysis, Big Data Tech, NLP

EXPERIENCE

Data Science Consultant – Unpaid   May 2023 – August 2023
Cook County Treasurer's Office – Sponsored by Illinois Tech (Python, PySpark, Jupyter Notebooks)   Chicago, IL
- Collaborated with domain experts, developed high performance classifier, increased predictive capabilities 100%
- Provided ability to send advanced warnings to targeted individuals potentially saving 90% on mailing expenses
- Ingested 6 years of public property tax data from Excel using PySpark to efficiently load and analyze data
- Preprocessed data by one-hot encoding, removing nulls or irrelevant features to improve data quality

Machine Learning Intern   May 2022 - August 2022
Motorola Solutions Inc., Applied Technology (Python, PyTorch, Bash)   Schaumburg, IL
- Implemented modifications to adapt Wav2Vec2 for Voice Activity Detection, aligning with project objectives
- Experimented training model with CNN and linear algorithms to identify optimal model for VAD
- Produced text file which classified every 20ms of speech output from trained model as speech or non-speech, simplified model scoring
- Documented detailed extraction to evaluation steps streamlining result reproduction
- Automated model experimentation and training processes by developing Bash scripts, enabling efficient iteration through various model architectures
- Presented key takeaways from the project to non-technical audience and its potential impact on business objectives

---

PROJECTS

Credit Card Fraud Detection Analysis (R, RStudio)
- Rebalanced training set using SMOTE in R increasing dataset size and improving data quality
- Led a team of 4 by delegating tasks, creating schedules and meetings, and producing deliverables all by deadline
- Improved accuracy of the logistic model and decision tree model by 12% and 6%, respectively
- Analyzed residuals and correlation to reduce feature size resulting in simpler and more efficient models

Employee Database (SQL, Java, DBeaver, PostgreSQL)
- Implemented a SQL script to create a database to store employee data for a simulated company
- Programmed a Java application to communicate with a PostgreSQL database to keep and update employee information
- Configured database to allow admin, manager, and employee users with different information privileges; improved privacy
- Managed a team of 3 by organizing meetings and delegating tasks to meet project deadline

---

SKILLS

Programming Languages: Python (NumPy, Pandas, Scikit-learn, PyTorch, PySpark), R (caret, dplyr), SQL, HQL, Java
Technologies: Jupyter Notebooks, Git, DBeaver, PostgreSQL, Linux/Bash, AWS (S3, EC2, EMR), Hadoop, Azure

---

EXTRACURRICULAR

Certifications: J.P. Morgan's SWE on Forage, Microsoft Learn AI Skills Challenge, Python for Data Science by Uplimit
Organizations: Member-Muslim Student Association, Member-Association of Computing Machinery, Member-ML Club
Volunteer: Guide at Sacred Learning Center during weekly Friday Prayer