Rahul Nair

rahulmnair1997@gmail.com • LinkedIn • Medium • Github • Portfolio

SUMMARY

Over 2 years of comprehensive experience in Data Analyzing, Business Intelligence, building Predictive/Forecasting models across various domains. Ability to solve complex business problems using ETL, Data Mining, Machine Learning & Data Warehousing concepts.

EDUCATION

- **MS in Data Science,** Illinois Institute of Technology, **GPA: 3.66** (Aug 2019 - May 2021)
  Coursework: Machine Learning, Big Data Technologies, Applied Statistics, Statistical Learning, Data Preparation and Analysis, Data Science Practicum
- **B.Tech in Computer Science,** University of Petroleum and Energy Studies, **GPA: 3.47** (Aug 2015 – May 2019)
  Coursework: Data Structures and Algorithm, Database Management and Warehousing, Information Retrieval, Discrete Mathematics, Computation and Automata

SKILLS

- **Programming:** SQL, Python, R, JAVA, C, C++, SAS, Pyspark, HTML, Agile Methodology, PostgreSQL, MySQL
- **Big Data Ecosystem**: Spark, Hadoop, MapReduce, Hive, Pig, Kafka, Flume, Hbase
- **Cloud Technologies**: AWS (S3, EC2, Lambda, Athena, RDS, Redshift, EMR, Sagemaker), NoSQL, Cassandra, MongoDB, Kubernetes, Google Data Studio, Microsoft Azure, Snowflake, CircleCI, Airflow, Prefect
- **Tools:** Tableau, Power BI, Azure ML, RStudio, Jupyter Notebook, SAS E-Miner, SAS CI, IBM-Unica, SSIS, MS Office, JIRA, Looker, GitHub, DBT, VS Code, DataGrip, Asana, Spyder, PyCharm
- **Libraries:** Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Keras, Nltk, Gensim, Scipy, Beautiful Soup, Tensorflow
- **Datasets:** HTTP, HTML, XML, JSON
- **Scripting:** Unix

WORK EXPERIENCE

Data Analytics Intern at M1 Finance (Jul 2021 – Present)

- Own and drive strategic analytical projects and insights to influence and support marketing initiatives. Develop and automate Weekly Business Reviews for stakeholders per week.
- Collaborate with stakeholders to define and operationalize **KPIs** for short and long-term measurements. Created **Tableau** dashboards to explain variation in success **Metrics** and **Time Series Analysis** to higher management.
- Develop and implement databases, data collection systems on **Redshift** using **DBT** and **version-control (Github)**, and other strategies to optimize statistical efficiency and quality.

Data Analyst (Practicum Student) at Labelmaster (Jan 2021 – May 2021)

- Optimized complex **SQL** scripts for quality checking of projects and populating output tables for deployment using **Azure Pipelines**.
- Automated hourly status report saving **10 man-hours/week**, thus decreasing response time for fixes and campaign failures.
- Achieved an accuracy of **MAPE 15%** approx. on price forecasting using **SARIMA**, further created web-app for presenting the forecasted values to the higher management using **HTML** and **CSS**.

Research Assistant at Illinois Institute of Technology (May 2020 – Oct 2020)

- Formulated ad-hoc reports based on requirements gathered from various stake holders using **JIRA** to provide solutions.
- Automated **ETL** processes using **Prefect** (Python), making it easier to wrangle data sets and reducing time by as much as 40% by performing large-scale data conversions, and transferring BAAN data into standardized formats.
- Developed interactive visualizations for stakeholders to interact with the data the way they please and create policies.

Machine Learning Engineer at Epic Minds IT Pvt. Ltd. (Jun 2018 – Jun 2019)

- Performed data collection of 18,000 images of diseased crops from various sources in collaboration with the Company Agriculture Research team.
- Built an end-to-end image classification model using **python** to predict the disease which the plant has with an **accuracy of 86%**. Trained **Google Inception V3** model for classification and used **Django** and **Flask** for web interface.
- Developed and automated **data migration pipeline** from SQL Server to Snowflake using **SnowSQL** and **SnowPipe**, and performed **dimensional modeling** on the migrated data, further created **data dictionary** for the technical audience.

PROJECTS

CLUSTERING AND REGRESSION ANALYSIS OF GERRYMANDERING .

- Implemented **weighted k-means** to develop a new redistricting plan for Pennsylvania making sure the population distribution remains proportionate across districts. This new plan has **improved the fairness by about 60%.**
- Experimented with certain regression analyses such as **Best Subset, Ridge and Lasso** to find which factors influence elections the most.

YELP RECOMMENDER SYSTEM FOR RESTAURANTS

- Developed an **end-to-end recommender system** leveraging **python** to suggest restaurants to users utilizing **hybrid matrix factorization** method with an **accuracy of 97%.** Deployed final model as a web-app employing **Angular JS** and **Flask**.
- Analyzed other algorithms such as **collaborative, content-based,** etc. and tuned the hyperparameters using **Bayesian**.