

# Abdul Naeem Shaik

+18623176286 | Hoboken, NJ | [ashaik13@stevens.edu](mailto:ashaik13@stevens.edu) | [Linkedin](#)

## Education:

- Masters in Data Science | Stevens Institute of Technology | Hoboken, New Jersey, USA** **September 2023 - May 2025**  
Coursework: Statistical Methods, Deep Learning, Machine Learning, Optimization, Time Series Analysis, Artificial Intelligence
- Bachelors in Electronics & Communication Engineering | GRIET | Hyderabad, India** **August 2018 - June 2022**  
Coursework: Python, Data Structures, Data Analytics, High Performing Computing, Data Mining, Database Management, AI/ML

## Professional Experience:

### Data Engineer | Nexus Software Limited | Hyderabad, India

June 2022 – August 2023

- Engineered and maintained scalable SQL databases (MySQL, Redshift), structuring over **1M+ records** to streamline cross-functional reporting
- Built and optimized **10+ automated ETL pipelines**, accelerating report delivery by **35%** and reducing manual tasks
- Refined raw datasets using Python and SQL, improving data accuracy by **40%** and enhancing reporting reliability
- Managed AWS infrastructure (S3, Redshift), achieving 99.9% uptime and cutting data retrieval time by 25%
- Collaborated with cross-functional stakeholders to implement data governance practices, reducing errors in pipeline outputs
- Documented data workflows improving handoff efficiency and process adoption by **30%**

### Data Analyst Intern | Vvn Technologies | Hyderabad, India

April 2021 – May 2022

- Analyzed geospatial datasets for **50+ districts**, optimizing public service delivery and resource allocation in government sectors
- Created **3+ interactive dashboards** to present location-based insights, improving access for **20+ stakeholders**
- Cleaned and standardized **100K+ records** using SQL and Pandas, improving dataset integrity by **35%**
- Integrated geospatial insights into a public portal, increasing citizen engagement by **40%** and reducing inquiry volume by **25%**

## Technical Skills:

- Programming Languages:** Python, SQL, R
- Frameworks & Libraries:** Pandas, NumPy, Scikit-learn, TensorFlow, Flask, PySpark, Airflow
- Databases & Warehousing:** MySQL, PostgreSQL, Snowflake, AWS Redshift
- Big Data & Cloud Technologies:** Apache Spark, Databricks, Azure Data Streams, AWS (S3, Glue)
- ETL & Data Pipelines:** Apache Airflow, dbt (Data Build Tool), SQL-based ETL
- Business Intelligence & Visualization:** Power BI, Tableau, Matplotlib

## Projects:

### Fraud Detection System | Apache Spark, Databricks, Azure Data Streams, Snowflake, Power BI, Python, SQL

- Designed a real-time fraud detection system processing **50K+ telecom calls/day**, boosting scam identification accuracy by **85%**
- Developed a K-Nearest Neighbors (KNN) model to detect fraudulent calls, storing outputs in Snowflake and Redshift for scalable querying and downstream analytics.
- Optimized complex SQL queries to accelerate fraud case investigations and ensure rapid insights into scam patterns.
- Implemented and maintained interactive Power BI dashboards for real-time fraud alerts, risk score tracking, and trend analysis, enhancing decision-making for fraud analysts.
- Automated ETL pipelines using Apache Airflow, eliminating manual processes and increasing fraud detection efficiency by 30%

### Statistical Modeling & Predictive Analysis of Lung Cancer Risk | Python, NumPy, Pandas, SciPy, Seaborn, Scikit-learn

- Analyzed **500K+ patient records** using Bayesian methods, identifying **5 key factors** contributing to **59%** of lung cancer cases
- Simulated multiple distributions (Normal, Poisson, etc.) and performed outlier detection using **Z-score and IQR**, improving statistical accuracy
- Enhanced data reliability by **30%** through statistical distribution testing and advanced outlier filtering
- Modeled Markov Chains to simulate patient health state transitions with 95% accuracy in predicting progression patterns
- Achieved 87% classification accuracy using a Naïve Bayes model to predict lung cancer risk from lifestyle and environmental factors.

## Certifications:

- IBM Data Science Professional Certificate**
- AWS Certified Data Analytics**
- Microsoft Power BI Data Analyst Associate**
- Databricks Lakehouse Fundamentals**