ROOPAM VERMA
Data Engineer| MSCS Graduate

E-Mail: roopamverma9717@gmail.com
Phone: (469)-885-2729

GitHub: roopamverma
LinkedIn: roopam-verma
Portfolio website: https://roopamrv.github.io/mywebsite

CAREER OBJECTIVE

Data Engineer with hands-on experience building scalable ETL/ELT pipelines using **Apache Spark, Databricks, and SQL** across **Azure and GCP**. Strong background in data modeling, batch processing, and data quality validation, with a focus on delivering reliable, analytics-ready datasets to support business decision-making.

EDUCATION

- *Master's degree in Computer Science - University of Texas, Arlington* [**Fall 2021 - May 2023**]
  [Key Courses: *Database Systems, Machine learning, Cloud Computing, Data Analysis and Modelling techniques, Data Mining, Theory of Computation, Operating Systems, Design and Analysis of Algorithm, Software Engineering, Artificial Intelligence, Web Data Management*]

- *Bachelor's Degree in Electronics and Communication Engineering – Banasthali Vidyapith, Rajasthan, India* [**July 2014 - May 2018**]
  [Key Courses: *Computer Architecture, Digital Communication, Object oriented Programming, Satellite Communication, Digital Electronics, Cpp Programming*]

SKILLS

Programming & Query Languages :	SQL, Spark SQL, Scala (Apache Spark), Python
Data Engineering & Big Data :	Databricks, Apache Spark, ETL/ELT Pipelines, Batch Processing, Data Integration, Data Quality, Data Modeling, Data Warehousing
Cloud Platforms :	Microsoft Azure, Google Cloud Platform (GCP), Amazon Web Services(AWS)
Databases & Storage :	Oracle, MySQL, BigQuery, MongoDB, Relational Databases, Cloud Object Storage
Tools & Platforms :	Informatica, Power BI, Git, Databricks Jobs, Jupyter Notebook, Visual Studio Code
Operating Systems :	Linux, Windows

WORK EXPERIENCE

Data Engineer Intern, Hudson Advisors [May 2023 – Present]

- Designed, developed, and supported scalable ETL/ELT and batch data pipelines using Oracle SQL, Informatica, Databricks, and Apache Spark (Scala/Spark SQL) for enterprise-scale structured datasets.
- Optimized Spark jobs, reducing pipeline runtime by ~30% and improving cluster resource utilization.
- Assisted in the migration and integration of on-premises data sources to cloud platforms, leveraging Databricks on Azure and GCP for scalable data processing and analytics.
- Leveraged cloud object storage, distributed compute clusters, and scheduled Databricks jobs to support scalable data engineering workflows.
- Wrote, optimized, and maintained complex SQL and Spark SQL queries for data transformation, validation, and analytics-ready datasets.
- Implemented data quality checks, validation logic, and error handling within ETL workflows to ensure data accuracy, consistency, and completeness.
- Developed and maintained data models (fact and dimension tables) to support downstream reporting, BI, and analytical use cases.
- Collaborated with cross-functional teams to translate business requirements into technical specifications, while maintaining clear documentation for data flows, transformations, and pipelines.

Systems Engineer, Infosys Pvt. Ltd., India [Oct 2018 - Feb 2020]

- Successfully developed and improved the Media Gateway Control Function product-GSX (Gateway Signalling Exchange) for Ribbon Communications, USA.
- Proficient in C++ programming language with a deep understanding of data structures, pointers, and memory management.
- Demonstrated expertise in SIP stack and sip-parser files for parsing new headers and fields in SIP and SDP messages.
- Knowledgeable in call flows of SIP, ISDN/ISUP calls, and protocols such as TCP/UDP, OSI layers, RTP, SMTP, DNS, etc.
- Proficient in Python scripting and utilized it to streamline work structure and perform various audits.
- Recognized for exceptional performance, awarded the "Winner of Hackathon" for developing an audit tool to stop Memory Leakage and enhancing the product's efficiency, resulting in a significant reduction in manual work from days to minutes.
- Experienced in working with LINUX IDE for C/C++ UNIX Shell Scripting and proficient in Python scripting and Database management using SQL.

Research Intern, Defense Research and Development Organization, Delhi, India [Jul 2017 - Dec 2017]

- Conducted research on Microstrip and Patch Antenna's functionality in Bluetooth (2.45GHz) and Wi-Fi (5GHz) frequency ranges.
- Developed a simulation using Application CST Microwave Studio and Proteus ISIS to study antenna performance and behavior.
- Investigated the role of Microstrip and Patch Antennas in mobile communication technology.

Intern Embedded Engineer, HCL Pvt. Ltd., Delhi, India [May 2016 - Jun 2016]

- Conducted research on Embedded Systems and gained expertise in the use of hexadecimal codes in Embedded C language.
- Demonstrated proficiency in Atmel Studio and ISIS Proteus Softwares for developing a RFID card Reader.
- Designed and executed simulations before implementing hardware components using program circuit board and integrated circuit ATMEGA 8.

TECHNICAL PROJECTS

Data Processing using Apache Spark
- Python, Scala scripts for data processing and presenting complex data, files based on the requirement (Click here for Work)

Machine Learning- Diabetes Severity Prediction, *University of Texas at Arlington*

- Designed ML models such as Logistic Regression, K-Nearest Neighbors, Random Forest, Support Vector Machine to understand and predict the extremity of diabetes of patients to reduce risk and made a webpage for user to input related field values and check the severity of diabetes. Accuracy ~ 84%

Student Discussion Board - Application Development using Python Flask and Deployment on Google Cloud Platform
- Developed Discussion board for UTA from the scratch using Python Flask and deployed it on Google cloud using App Engine, CloudBuild and CloudSQL and created an event driven pipeline for CI/CD using GitHub. (Click here to view work)

Auto-Insurance Customer Analysis & Prediction using K-means Clustering, *University of Texas at Arlington*
- Designed an ML model that predicts changing Auto insurance based on its historical prices over a period of five years, given the respective news articles post-closing time of the previous day and other globalization factors. Employed Machine learning and Deep learning algorithms such as K-means Clustering and Hierarchical clustering to understand and predict the vehicle insurance. Achieved ~ 83% accuracy.

Analysis of Supervised Learning Models to predict Diabetes severity, *University of Texas at Arlington*
- Designed ML models such as Logistic Regression, K-Nearest Neighbours, Random Forest, Support Vector Machine to understand and predict the extremity of
  diabetes of patients to reduce risk and made a webpage for user to input related field values and check the severity of diabetes.
  Accuracy ~ 84% (Click here to view blog)

CIFAR-10 image classifier using Convolutional Neural networks, *University of Texas at Arlington*
- Built a Neural network model with multiple convolutional layers (CNN), cross-entropy function and SGC optimizer. Improved the performance by increasing number of layers and epochs. Accuracy ~ 65% (Click here to view blog)

Naïve Bayesian Classifier used for sentiment analysis of Imdb reviews, *University of Texas at Arlington*
- Built a Naive Bayesian model that processes large amounts of Imdb reviews data and classifies them into positive and negative reviews based on the words used and their prior/posterior probabilities. Applied Laplace smoothing and studied the efficiency variations. (Click here to view blog)

Web Application on Attendance Tracking System, *Banasthali Vidyapith, Rajasthan, India*
- Designed and developed website for attendance tracking with features including Attendance management as well as viewing whole day schedule for faculty as well as students, reducing the paperwork and making it time efficient.

CLUBS AND CERTIFICATION
- **Outreach Lead** – Innovacation Club Core Team at Banasthali Vidyapith.
- **UDEMY Certificate –** Python for Machine Learning and Data Science
- **Volunteer** – National Service Scheme active member worked to ensure that everyone who is needy gets help to enhance their standard of living and how to lead a good life despite a scarcity of resources.

ACHIEVEMENTS
- Winner of Hackathon at Infosys for adding an innovative feature to existing Media Gateway
- First position in 'Robot Race' a technical event held by innovacation club at Banasthali Vidyapith
- First position in the event 'Phyction' in the National Technical Fest 'Mayukh 2k16' held at Banasthali Vidyapith
- First position in the event 'Divertido-Twiddle' an interview event conducted by ICE Club held at Banasthali Vidyapith
- Second Position in Model Making Competition (Working Van-De-Graff Generator) held at School
- Member of Student council as House Leader at School.
- Winning position at District Level Race competitions