

Jaskirat Singh

716-563-5994 | jaskirat0810@gmail.com | [linkedin: jaskirat-singh-pahwa](https://www.linkedin.com/in/jaskirat-singh-pahwa) | [github: jaskirat-singh-pahwa](https://github.com/jaskirat-singh-pahwa) | NY, USA

EDUCATION

University at Buffalo – SUNY

Master of Science in Data Science (GPA - 3.90/4.0)

Jan 2023 – May 2024

Buffalo, New York

Guru Gobind Singh Indraprastha University

Bachelor of Technology, Computer Science and Engineering (CGPA - 8.4/10.0)

Aug 2014 – May 2018

New Delhi, India

TECHNICAL SKILLS

Areas of Expertise: Data Warehouse Modeling, Data Analytics, Business Intelligence, Databases, Machine Learning

Languages and Databases: Python, SQL, Scala, R, Java, Bash, MySQL, PostgreSQL, NoSQL (MongoDB)

Big Data: Apache Spark, Hadoop, Airflow, Kafka, Snowflake, Superset, Tableau, ETL, DBT, Databricks

Cloud - AWS: Athena, Glue, Step Functions, Lambda, S3, Sagemaker, EC2, EMR, SNS, Redshift, Cloudwatch, IAM

Libraries: PySpark, Boto3, Sklearn, Pandas, Deequ, PyTorch, NLTK, NumPy, Matplotlib, Keras

Tools: CI/CD, Docker, Terraform, IaC, Agile (Scrum / Kanban), Git, DevOps, GitLab, Jenkins, Jupyter, JIRA, Linux

WORK EXPERIENCE (4.5 YEARS)

Machine Learning Engineer, Tiger Analytics

Oct 2021 – Dec 2022

Platform, Data, and ML Engineering

Remote - Bangalore, India

- Collaborated with 4 domain experts to create ML models for Lung Cancer detection and stage prediction
- Engineered over 250 features from large-scale clinical datasets using PySpark, S3, and Glue jobs
- Productionalize end-to-end Machine Learning workflows using Sagemaker pipelines
- Transitioned data pipelines from Step Functions to Airflow, expanded integrations beyond AWS
- Addressed compliance issues for 4 AWS services, improved security using Boto3, and Lambda

Data Engineer, Sahaj Software Solutions

July 2019 – Sep 2021

Data Science and Engineering

Bangalore, India

- Implemented data-centric solution for optimizing out-of-home ads, using geospatial data for real-time analytics
- Deployed ETL processes to ingest data from 4 vendors into S3 which enhanced accessibility by 30%
- Orchestrated 10+ data pipelines and developed algorithms to generate audience insights and impressions
- Scaled platform to handle 500+ GBs of data, managed 10+ billion observations every week
- Discovered and resolved 3 crucial data quality issues using Deequ, fostered credibility and predictive accuracy
- Built 5+ impactful Superset dashboards using SQL and Athena, elevated campaign planning for US & UK region
- Tailored RASA chatbot for top Indian transport firm, refined core NLP features - NEL and intent ranking

Trainee Software Developer - R&D, QOS Technology

July 2018 – June 2019

Data Analysis, Machine Learning, and Deep Learning

Bangalore, India

- Designed a system to block IP addresses/domains on Checkpoint Firewall based on score threshold
- Analyzed unstructured real-time firewall logs, and created 2 attributes to determine severity levels
- Developed a Neural Network to calculate risk score of incident logs using Python and TensorFlow
- Automated repetitive SOC analysts' tasks using APIs in Python, leading to 5x faster incident response time

PERSONAL PROJECTS

Data-Modelling for Airbnb Analytics - [Link](#) | *Python, Pandas, PostgreSQL, Tableau*

- Transformed raw flat file data into an optimized star schema for data analysis, increasing efficiency and scalability
- Constructed advanced SQL queries for insights; built a Tableau dashboard for interactive data visualization

Analyzing Residential Real Estate Financing Patterns - [Link](#) | *R, RegARIMA, VAR, LSTM*

- Analyzed relationship between US real estate loans and mortgage rates using ARIMA, VAR, and LSTM models
- Achieved accurate forecasting of loan volumes, providing valuable insights for lending strategies and market trends

Sentiment Analysis - IMDB 50k Movie Reviews - [Link](#) | *Python, NumPy, PyTorch, RNN, Gensim, LDA, NLP*

- Executed pre-processing tasks - text cleaning, word-to-vector conversion, and sentiment dictionary creation
- Fine-tuned a Recurrent Neural Network on 40,000 sentiments, attaining 90% prediction accuracy on test data

Used Cars Price Prediction - Scraped 370k Observations - [Link](#) | *Python, Pandas, Scikit-Learn, Streamlit*

- Conducted EDA, data cleaning, extracted new features, modeled 10 regression algorithms, and achieved R2 of 0.88
- Built a Streamlit web app for visualizing market trends on popular car brands for target buyers