

# Swarup Das

## Team Lead | MSc. Data Science | Machine Learning | MLOps | Data Engineering

Pune, India • +91 9503554884 • [dasswarup53@gmail.com](mailto:dasswarup53@gmail.com) • <https://www.linkedin.com/in/swarupd/>

Data and ML lead with six years experience in building and productionizing machine learning models and ETL pipelines at scale. My forte resides in building scalable & machine learning powered end to end products.

### WORK EXPERIENCE

---

#### Lead Machine Learning Engineer (Fintech)

March 2020 – Present

OneCard by FPL Technologies, Pune, India

Built amazing products like OneScore (20 Million + Users) & OneCard (~ 2 Million Users). Startup gained its unicorn tag in just 3 years. Led the team of engineers to create Data Lake, ML Workbenches & Analytics Platform.

- Trained and containerised a real time fraud detection system using top notch regression strategies that mitigated frauds by 98%
- Built a dynamic offer recommendation engine using location based intelligence and real time data ingestion pipelines.
- Tuned BERT, LLAMA2 (entity extraction & sentiment analysis) to improve CX experience and boost first call resolution by 45%. (currently exploring other LLMs as well)
- Reduced signup drop-offs from 65% to 15% and increased user engagement by 40%, through a combination of hypothesis testing, segmentation analysis and A/B testing.
- Developed and deployed models using Ensemble Learning & Feature Engineering methods that powers Risk, Demand Forecasting, Onboarding and Recommendation Engines
- Penned highly efficient time series models using SARIMAX, RNNs for spends projection, payment defaults and anomaly/outlier detection.
- Engineered an ML Platform that was featured by AWS. ([Click Here](#)).
- Developed POCs for inhouse projects using Deep Learning, Clustering, Natural Language Processing and Computer Vision (YoloV8, DeepFace).
- Penned & Owned AWS MLOps pipelines using Docker, AirFlow, MIFlow, GitOps etc. that automatically trains, tunes and deploys 120+ risk, credit scoring, churn and LTV models.
- Programmed highly scalable and secure Flask Api pipelines for taking ML models to production environments.
- Crafted several dashboards for model performance tracking, governance and drift monitoring using business critical and statistical metrics along with tools like Prometheus and Grafana.
- Mastered complex SQL query development and performance tuning/optimization using partitions
- Solved the resource starvation problem by implementing On Demand Containerization.
- Leading the migration from AWS Data Lake to Delta Lake using databricks (125s+ TBs of Data).
- Automated the model maintenance and monitoring using AWS Sagemaker to tackle data distribution drift.
- Led the first major effort to build the serverless data lake that served as the base of the analytics platform and it also improved the operational efficiency by 35%
- Crafted highly scalable AWS Glue ETL pipelines using Hive, Hudi & Spark that updates 1200+ SQL tables on a daily basis which revolutionized the entire data warehouse stack.
- Performed R&D on data storage systems and compression techniques using Delta, Hudi, Iceberg formats to optimize the storage and querying costs that overall reduced the operational cost by 45%.
- Built the governance and auditing module that is compliant with partner banks, on top of our existing data infrastructure to ensure data consistency, reliability and availability.
- Collaborated closely with CXO's, data analysts and domain experts to build highly scalable & secure data products to tackle challenging business problems.
- Mentoring & grooming data analysts for technical queries and architecting complex ML pipelines.

## Data Scientist

Jan 2019 – Feb 2020

Rule 14 LLC, Pune, India

First hire of the data science team and was also responsible for team building and heading initial client projects

- Worked on Text Summarisation using Encoder Decoder models, forecasting using ARIMA/SARIMAX
- Developed a variety of in house projects using PCA, XGboost, Random Forest, Text Rank algorithms

## Software Engineer Trainee

July 2018 – Dec 2018

Searce LLC, Pune, India

- Worked on Image captioning (using VGG16), Model Blending and learnt best practices like CI/CD, Canary Deployment, Test Driven Development etc

## EDUCATION

---

- **MSc Data Science (Year 2021)**, Liverpool John Moores University, Grade : Distinction
- **Bachelor of Engineering Computer (Year 2018)**, Pune University, Grade : Distinction

## SKILLS & OTHER

---

**ML Algorithms:** RandomForest, XGBoost, K-Means, SVM, Naive Bayes, PCA, Linear & Logistic Regression,

**Libraries:** Keras, PyTorch, Tensorflow, Airflow, Sklearn, Seaborn, Matplotlib, Numpy, Pandas, Sanic, Flask, FastAPI, Pyspark, Git

**Big Data/DataBases:** Hive, Hudi, Spark, IceBerg, Kafka, MySql, PostgreSQL, BigQuery, DataBricks

**Cloud:** AWS EC2, RDS, DMS, Lambda, Glue, Studio, Event Bridge, Sagemaker, Athena, Kinesis, S3, ECR, Docker, Kubernetes

**Languages:** Python, SQL, Java