Name: Dushyant Singh

Risk and Decision Scientist

A motivated team player with a results-oriented mindset, excelling in Data Science and Risk Analytics with over 4 years of experience. A quick learner who delivers results in minimal time while maintaining effective communication. Passionate about applying data analytics to solve real-world problems.

Email: dushyant.singh.civ16@iitbhu.ac.in
Contact: 7014093420
City: Pune, India

Linkedin: linkedin.com/in/dushyant-singh- 3214a8144

Github: github.com/dushyant4342

Risk Data Scientist & Data Management 03/2024 - Present

A results-driven Data Scientist with over 4 years of experience in Risk Analytics, Data Management, and Machine Learning. Skilled in developing data-driven solutions that optimize risk assessment, automate workflows, and enhance operational efficiency. Adept at collaborating across teams, implementing scalable data pipelines, and leveraging advanced analytics for business impact. Quick learner with a strong problem-solving mindset, ensuring timely execution of complex tasks.

SKILLS

AWS (Lambda, Glue, Kinesis, Firehose, S3, EventBridge, Athena, SageMaker) | Python | SQL | Machine Learning | Neural Networks | NLP | OpenMetadata | Superset | Data Engineering | Risk Modeling | Statistics | Leadership | MS Office | SPSS | LangChain | EC2 | LLM WebApp | Docker |

SKILLS

Cloud & Data Engineering: AWS (Lambda, Glue, Kinesis, Firehose, S3, EventBridge, Athena, SageMaker, EC2) | Docker | Big Data Processing (PySpark, Trino, Delta Lake, Polars) | ETL & Data Pipeline Optimization

Programming & Analytics: Python | SQL | Machine Learning | Neural Networks | NLP | Statistics | Risk Modeling | XGBoost | Decision Trees | Random Forests | Linear Regression

MLOps & AI: MLOps (MLflow, Feature Store, KSI/PSI Tracking) | AI Model Deployment (DeepSeek, LLaMA 3, AWS Bedrock, Ollama, LangChain) | Retrieval-Augmented Generation (RAG)

Data Governance & BI: OpenMetadata | Superset | Data Governance & Quality (PII Masking, Data Lineage) | Business Intelligence (GChat Alerts)

Tools & Frameworks: GitLab CI/CD | Flask | Streamlit | SPSS | MS Office | LangChain | pdfplumber

WORK EXPERIENCE

FPL Technologies (OneCard) - Pune
03/2024 - Present
Role: Lead Data Management
Led the end-to-end data migration by organizing tables bank-wise into structured L1/L2/L3 formats. Developed AWS Glue Jobs (PySpark) to migrate data into Delta format tables, streamlining ETL pipelines. Consolidated raw tables by incorporating required columns based on collected requirements, eliminating the need for multiple table joins in raw tables and enhancing efficiency.
Developed CI/CD pipelines for cron and (access creation) instance migrations using Starburst (Trino), S3 sync with GitLab, and Lambda functions for rule creation. Guided the Data team of 80+ people for better data management, enforce query optimizations and cost cutting by shifting workloads on Polars and PySpark.
Led the AWS Feature Store project with Lambda-triggered pipelines for seamless feature sharing across teams. Implemented MLflow for model monitoring and governance, tracking stability metrics (KSI, PSI) on Superset without manual intervention and reporting discrepancies on Google Chat.
PII data Security: Wrote Glue jobs for identification of PII tables, implemented masking with Ranger to ensure hash-based joins. Built Lambda-driven & Gitlab workflows for executing Jupyter scripts, Glue jobs, SFTP uploads, and API calls, ensuring safely PII data processing.
Moved Superset to Trino & EC2 for cost optimisation, upgraded it to the latest version and enhanced functionality with PDF exports, new charts and GChat notifications. Also, Integrated OpenMetadata for data dictionary, lineage tracking, test cases, profiler for data ingestion and email alerts while conducting training sessions for all the team.
Built real-time data streaming pipelines using Amazon Kinesis, Firehose and EventBridge for event routing. Integrated a Lambda function within Firehose to preprocess streaming data by extracting relevant keys from callback tables before storing them in destination table.
Developed a centralised warning system to track key business metrics (e.g., cards, KYC success%, repayments). Also, worked on root cause analysis using the DoWhy library to identify the anomalies.

Sabbatical - 6 Months

FPL Technologies (OneCard) - Pune
10/2020 - 08/2023 (2.11 years)
Role: Risk and Decision Scientist

- Developed an XGBoost model for risk segmentation (1M customers), achieving 81% Gini and improving resolution rates from 89% to 92.4%.
- Designed risk triggers, EMI, Autopay, and campaign strategies that reduced early delinquency rates from 3.18% to 2.54% on a 3000+ Cr portfolio.
- Created a propensity-to-pay model for Bucket-1 customers, improving resolution rates by 13% on a ~145Cr portfolio.
- Automated multiple manual workflows on AWS (agency allocation, repayments, dashboards, emails, FLDG, model variables), enhancing efficiency.
- Built a Random Forest Classifier for optimizing communication channels, reducing spamming, and increasing NPS by 7%.
- Developed a Decision Tree model for agency allocation in Bucket 2-3, reducing flow into NPA, lowering losses by 9%, and cutting POF by 0.23%.
- Collaborated with the sourcing team to optimize credit risk strategies, leading to fraud detection and lower credit losses.
- Led risk-based communication strategies, improving DIY repayments by 3% and increasing LPC revenue by ~30%.

- Led the development of OneAccord, an internal collection system integrated with dialers, nudges, billing, and contact databases, boosting data security and productivity.

COVER LETTER

I am excited to apply for Data **Science**. With expertise in **data analytics, machine learning, ETL Jobs, data pipelines and product development**, I bring a unique blend of **data management and data science** experience that optimizes both efficiency and business decision-making.

At **FPL Technologies (OneCard)**, I led the end-to-end **data migration**, organizing tables into structured **L1/L2/L3 formats** to improve accessibility and performance. I developed **AWS Glue Jobs (PySpark)** to migrate data into **Delta format tables**, automating ETL pipelines and **reducing compute costs**. Additionally, I built **CI/CD pipelines** for cron and instance migrations using **Starburst (Trino), S3 sync with GitLab**, and Lambda functions for rule management, significantly streamlining data infrastructure. Security and governance were at the forefront of my work, where I implemented **PII data security measures** such as **hash-based masking using Ranger**, automated PII detection with Glue jobs, and Lambda-driven workflows for **secure SFTP uploads, API calls, and Jupyter executions**. To improve system observability, I migrated **Superset to Trino & EC2**, **optimized costs**, upgraded it to the latest version, and introduced **PDF exports, GChat alerts, and data quality, data dictionary via OpenMetadata**.

Beyond data infrastructure, I have worked extensively on **real-time data streaming** by integrating **Amazon Kinesis, Firehose, and EventBridge** for efficient event routing. I also developed a **Lambda-based Firehose transformation** to extract and preprocess callback table data, ensuring high-quality real-time ingestion.

On the **data science** side, I built **ML-driven risk segmentation models using XGBoost**, achieving **81% Gini**, which helped in identifying high-risk customers and optimizing collection strategies. I developed **Random Forest Classifier models** for predicting the **best communication channels**, reducing spam while increasing engagement, and improving **Net Promoter Score (NPS) by 7%**.

Additionally, I worked on **propensity-to-pay modeling** for delinquent customers, achieving a **72% Gini** score, which allowed for **optimized agency allocation** and led to a **13% improvement in repayment resolution**. By developing **decision-tree-based repayment prediction models**, I enabled more accurate **stab-rollback logic for delinquent accounts**, resulting in a **9% reduction in losses and a 0.23% decrease in POF**.

I played a key role in **developing risk strategies and optimizing collection processes**. I designed **risk triggers, EMI, Autopay, and Jackpot campaigns**, alongside **improving calling scripts and communication journeys**, which led to a **reduction in entry rates from 3.18% to 2.54% on a 3000+ Cr portfolio**, saving over **20+ Cr in potential flow**.

I automated **manual workflows on AWS**, including **agency allocation, repayment and communication dashboards, payout emails, FLDG processes, and model variable tracking**, significantly **reducing manual intervention and increasing operational efficiency**.

Leveraging **ML-driven insights**, I built a **Decision Tree model** to predict repayment capability for **Bucket 2-3 customers**, applying it to **agency allocation and stab-rollback logic**, which successfully **reduced NPA flow, lowered losses by 9%, and decreased POF by 0.23%**.

Beyond modeling, I partnered with the **sourcing team** to **analyze credit risk across different cohorts**, taking strategic actions to **minimize losses through bank migrations, new city launches, bureau delinquency analysis, and fraud detection via high-risk patterns**. Additionally, I optimized **billing cycles and risk-based communication strategies for five new banks**, achieving **consistent entry rates of 2.6%, increasing LPC revenue by 30%, and improving DIY collections by 3%**.

Finally, as a **product manager**, I collaborated with multiple teams to develop **OneAccord**, an internal **collection system**, seamlessly integrating it with **Dialer, Nudge, Billing, and contact databases**, ensuring **data security and boosting team productivity**.
My ability to **bridge data science with data Engineering** & Devops has allowed teams to operate more efficiently while maintaining strong governance and security.

Technical Skills

- **CI/CD & Automation** (GitLab CI/CD, Lambda-triggered workflows, Jupyter automation)
- **Data Governance & Security** (OpenMetadata, PII masking with Ranger, hash-based joins)
- **Real-time Data Processing** (Kinesis, Firehose, EventBridge, Lambda transformations)
- **Big Data Processing & Query Optimization** (BigQuery, Trino, Delta Tables, Polars, PySpark)
- **Risk Analytics & Predictive Modeling** (XGBoost, Decision Trees, Random Forests, Linear Regression, NLP, ANN)
- **Product & Strategy Development** (Risk triggers, collection optimization, agency allocation models)
- **Business Intelligence & Observability** (Superset, Google Chat alerts, data dictionary, lineage tracking)
- **Cloud Cost Optimization** (EC2 migration, cost-cutting strategies with PySpark & Polars)
- **AI Model Deployment & Serving** (DeepSeek-R1, LLaMA 3, AWS Bedrock, Ollama, LangChain)
- **Retrieval-Augmented Generation (RAG)** (DeepSeek, LangChain, pdfplumber for document processing)
- **Containerization & Model Packaging** (Docker, Docker Hub, EC2-hosted AI models)
- **MLOps & Model Monitoring** (MLflow, Feature Store, KSI/PSI tracking)

Advanced Technical Skills

- **Data Architecture & Management** (L1/L2/L3 Data Structuring, Delta Lake)
- **ETL & Data Pipeline Optimization** (AWS Glue, PySpark, Trino, Polars)
- **Streaming Data Processing** (Kinesis, Firehose, EventBridge, Lambda Transformations)
- **MLOps & Model Deployment** (MLflow, Feature Store, Model Stability Metrics - KSI, PSI)
- **Data Governance & Compliance** (OpenMetadata, Ranger, PII Masking, Data Lineage)
- **CI/CD & Infrastructure Automation** (GitLab CI/CD, Lambda-Driven Workflows)
- **Decision Science & Risk Modeling** (Propensity Models, Risk Triggers, Agency Allocation Models)
- **API Development & Integration** (Flask, Streamlit, FastAPI, SFTP, REST APIs, Event-Driven Architectures)
- **Cost Optimization & Resource Management** (EC2 Migration, AWS Cost Efficiency Strategies)
- **Business Intelligence & Reporting** (Superset, Dashboard Automation, GChat Alerts)
- **AI Application Development** (Deployed DeepSeek & LLaMA 3 models on EC2 using Docker, Streamlit, Langchain)
- **GitLab on EC2** (Configured GitLab Runner for automated repository synchronization with AWS S3)

Soft Skills & Leadership

- **Team Leadership & Mentorship** (Guiding 80+ Data Professionals, Training on OpenMetadata & Superset)

- **Cross-Functional Collaboration** (Risk, Engineering, and Product Teams)
- **Problem-Solving & Strategic Thinking** (Root Cause Analysis with DoWhy, Risk Strategies)
- **Process Optimization & Automation** (Reducing Manual Workflows, Enhancing Operational Efficiency)
- **Stakeholder Communication & Data Storytelling** (Translating Data Insights into Business Decisions)

ACHIEVEMENTS

One Accord Award - FPL Technologies

Successfully completed the OneAccord project within two-months, closely coordinating with team members and detailed planning.

Honorable Mention - Gymkhana Award

Contribution in the growth of Film and Media Council at IIT BHU Varanasi

PERSONAL PROJECTS

Deployed AWS Bedrock model locally used LangChain, Bedrock client API and Streamlit ("amazon.titan-embed-text-v2:0" & "meta.llama3-8b-instruct-v1:0"). Also, explored "amazon.titan-image-generator-v1" for image generation.

Deployed LLama3(meta.llama3-8b-instruct-v1:0) model on ec2 using streamlit, github and docker container for Text summarization & classification (http://13.201.81.185:8501/)

Deployed "deepseekr1:1.5b" model for python code debug on ec2 using streamlit, github and docker container.

DeepSeek Model Deployment on EC2
Deployed a DeepSeek-based AI application utilizing the DeepSeek R1:1.5B model, managed via GitHub for version control. Implemented a Dockerized environment incorporating Ollama, LangChain, and Streamlit, and successfully deployed the application on an Amazon Linux 2023 EC2 instance.
End To End RAG Agent With DeepSeek-R1 And Ollama (Locally) (streamlit run rag_app.py) - read from pdf using pdfplumber
Deployed GitLab on an EC2 instance, configured GitLab Runner for CI/CD pipelines to synchronize repositories with AWS S3, and implemented AWS Lambda functions triggered to schedule on EventBridge and execute Jupyter notebooks.
I built and deployed a DeepSeek-based AI application using the **DeepSeek R1:1.5B** model. Initially, I deployed it locally using **Ollama** for model management, **Streamlit** for the frontend, and **LangChain** for AI interactions.

Development & Deployment Steps:

1. **Local Setup:**
    - Created a GitHub repository to manage the project.
    - Built a **Dockerfile** with all dependencies: **Ollama, LangChain, DeepSeek, and Streamlit**.
2. **EC2 Deployment:**
    - Launched an **Amazon Linux 2023** EC2 instance.
    - Installed **Git** and cloned the repository.
    - Installed **Docker** and used the **Dockerfile** to build an image.
    - Ran the container to deploy the application on EC2.
    - Allowed Inbound web traffic to 8501 port from any ip and accessible on web url publicip:port
3. **Containerization & Version Control:**
    - Uploaded the **Docker image** to **Docker Hub** for easy versioning and future reuse.

This setup ensures a scalable and reproducible AI-powered web app, leveraging Docker and AWS EC2 for efficient deployment.

House Price Prediction (05/2020 - 06/2020)

Performed data preprocessing, EDA, feature engineering and feature selection.

Utilized multiple algorithms and conducted performance comparisons to select the best model.

Fake News Classifier - RNN LSTM (07/2020 - 08/2020)

Built a Deep Learning model using NLP Techniques like Stopwords, PorterStemmer, Onehot, Embedding and RNN-LSTM to identify fake news (91% accuracy)

Got exposure to NLP techniques like Bag of words, TFIDF, n-grams and Word2Vec.

Artificial Neural Network (08/2020 - 09/2020)

Churn Modelling to predict exit of a customer from the organization.

Used Feature Scaling and Grid Search CV to identify best parameters.

EDUCATION - Indian Institute of Technology BHU

B-Tech in Civil Engineering (07/2016 - 04/2020)

INTERESTS
Photography
Drone-Cinematography
Swimming