

# Md Jahid Hasan

📍 Oxon Hill, MD 20745 📞 (240)-273-6070 🌐 [msjahid.me](https://msjahid.me) ✉ [msjahid@msjahid.me](mailto:msjahid@msjahid.me) 🌐 [msjahid](https://www.linkedin.com/in/msjahid) 🌐 [msjahid](https://github.com/msjahid)

**Kaggle Grandmaster** | Top 0.1% globally · Expert → Grandmaster in 6 months · 20+ highly-ranked notebooks

## Professional Summary

Data Scientist and ML Engineer with **6+ years** of professional experience building predictive models, scalable data pipelines, and production-ready ML systems. Proficient across the full ML lifecycle from feature engineering and model training through to REST API deployment and monitoring. Completing an MS in Data Science at Eastern University (GPA: **4.00/4.00**, expected May 2026). Delivered measurable results, including a **60%** reduction in report generation time, **85%+** model accuracy on recommendation systems, and pipelines processing **500K+** records daily with **99.9%** accuracy. Proficient in Python, R, SQL, TensorFlow, and PyTorch.

## Education

**Eastern University** Jan 2024 – Expected May 2026  
*Master of Science (MS), Data Science (GPA: 4.00 / 4.00)* St. Davids, Pennsylvania

- **Relevant Coursework:** Introduction to Statistical Modeling, Data Analytics in R, Data Manipulation, Applied Machine Learning, Natural Language Processing, Data and Database Management with SQL

**Southeast University** Sep 2013 – Dec 2017  
*Bachelor of Science (BS), Computer Science and Engineering (GPA: 3.78 / 4.00)* Dhaka, Bangladesh

- **Relevant Coursework:** Database Design, Artificial Intelligence, Statistical Methods & Probability, Image Processing, Data Mining, Linear Algebra and Matrices, Web and Internet Programming

## Experience

**Ludwig Pfeiffer Hoch- und Tiefbau GmbH & Co. KG** Jan 2021 – Dec 2023  
*Data Scientist* Dhaka, Bangladesh

- Developed and maintained robust database management systems handling **100K+** records using MySQL and PostgreSQL, ensuring data integrity and optimized query performance.
- Created and maintained automated reporting pipelines using Python (Pandas), SQL, and Power BI, reducing report generation time by **60%** and improving operational visibility.
- Built and deployed production-grade REST APIs using Django REST Framework, serving ML model predictions with low-latency inference and integrating seamlessly with frontend and enterprise systems.
- Implemented Microsoft Power Apps and Power Automate solutions to streamline and automate business workflows across departments, improving process efficiency by **40%**.
- Designed and implemented computer vision and ML models for automated pipeline inspection, reducing manual inspection overhead and improving defect detection accuracy.

**TAPPWARE Solutions Limited** Sep 2019 – Dec 2020  
*Software Engineer* Dhaka, Bangladesh

- Designed scalable ETL data pipelines for e-governance platforms, processing **500K+ records daily** with **99.9%** accuracy, ensuring reliable data flow across government and enterprise systems.
- Built and deployed scalable machine learning web APIs using Django REST Framework and PostgreSQL, supporting real-time inference for government and enterprise applications.
- Built personalized recommendation models using TensorFlow and Scikit-learn, achieving **85% accuracy** in predicting learner preferences and improving e-learning engagement across government platforms.
- Conducted advanced data analysis on Japanese patient datasets using statistical modeling and NLP techniques, extracting meaningful patterns to support data-driven clinical decisions.

**Qtec Solution Limited** Jan 2018 – Aug 2019  
*Junior Software Engineer* Dhaka, Bangladesh

- Applied a broad range of machine learning techniques, including classification, regression, clustering, and ensemble methods, to generate actionable business insights from large-scale datasets.
- Developed and fine-tuned sentiment analysis models for marketing strategy optimization, achieving up to **82% accuracy** and delivering actionable insights to support campaign decision-making.
- Built automated web scraping solutions and end-to-end data preprocessing pipelines using Python and BeautifulSoup to collect, clean, and analyze market trends and customer behavior at scale.
- Designed and implemented data warehousing solutions using dimensional modeling and ETL processes, enabling efficient storage, retrieval, and reporting across large-scale business datasets.

## Projects

---

### Customer Profitability & Marketing Analysis | R, Python, ggplot2, Matplotlib, Seaborn, RMarkdown, HTML/CSS

- Conducted comprehensive customer behavior and profitability analysis on **30K+** records using R and Python, examining CLV, marketing channel effectiveness, and profit drivers across **13 variables**.
- Created interactive ggplot2 visualizations with custom Rosé Pine themes to identify marketing channel performance trends, customer segmentation patterns, and outlier behavior across demographics.
- Delivered actionable insights through RMarkdown reports, including outlier analysis using IQR, Z-Score, and Modified Z-Score methods to support marketing strategy and customer retention.

### Customer Insights and Transaction Trends Dashboard | Python, Pandas, NumPy, Matplotlib, Seaborn, Time Series

- Designed an interactive analytics dashboard analyzing **100K+** financial transactions across **41 countries**, uncovering spending patterns by gender, membership status, payment method, and service usage.
- Applied univariate, bivariate, correlation, and time series analysis using Python, Pandas, and Seaborn to surface trends across **5** payment methods, **4** membership tiers, and transaction patterns over time.

### Multi-Language Detection Web Application | Python, Flask, Scikit-learn, Docker, Railway, HTML/CSS/JavaScript

- Developed and deployed an ML-powered web application capable of detecting **17 languages** with **98% accuracy**, training a Gaussian Naive Bayes classifier on a multilingual text dataset with preprocessing.
- Deployed a full-stack production-ready solution using Flask, Docker, and Railway with a REST API backend, enabling scalable, low-latency real-time language inference accessible via a clean web interface.

## Publications

---

### SmolLab SEU at CheckThat! 2025: How well do multilingual transformers transfer across news domains for cross-lingual subjectivity detection

CLEF - Conference and Labs of the Evaluation Forum, Sep 26, 2025

## Technical Skills

---

**Programming Languages:** Python, R, SQL, Java, C++, MATLAB, JavaScript

**Machine Learning:** Scikit-learn, XGBoost, LightGBM, CatBoost, Random Forest, Gradient Boosting, SVM, Ensemble Methods, Feature Engineering, Model Evaluation, Hyperparameter Tuning, Cross-Validation

**Deep Learning & Computer Vision:** TensorFlow, PyTorch, Keras, CNNs, RNNs, LSTMs, Transformers, Transfer Learning, OpenCV, Image Processing, Object Detection, Image Classification, Data Augmentation

**NLP & Text Analytics:** NLTK, spaCy, Hugging Face Transformers, Word Embeddings (Word2Vec, GloVe), Sentiment Analysis, Text Classification, Named Entity Recognition (NER), Tokenization, Topic Modeling

**Data Science Libraries:** Pandas, NumPy, SciPy, Statsmodels, Scikit-learn, Dask, Polars, PySpark

**Data Visualization:** Matplotlib, Seaborn, Plotly, ggplot2, Tableau, Power BI, Streamlit, Dash, Bokeh, Altair

**Statistical Analysis:** Hypothesis Testing, A/B Testing, Regression Analysis, Time Series Analysis, Experimental Design, ANOVA, Correlation Analysis, Statistical Modeling, Feature Selection, Dimensionality Reduction (PCA)

**Databases:** MySQL, PostgreSQL, MongoDB, SQLite, Database Design, Query Optimization

**Big Data & Cloud:** Apache Spark, AWS (S3, EC2), Docker, Kubernetes, ETL Pipelines, Data Warehousing

**Web & APIs:** Flask, Django, FastAPI, REST APIs, HTML/CSS, JavaScript, Node.js, BeautifulSoup (Web Scraping)

**Tools & Platforms:** Git, GitHub, Jupyter, Linux, VS Code, Postman, Swagger

**Microsoft Power Platform:** Power BI, Power Apps, Power Automate, SharePoint

**Methodologies:** Agile, Scrum, CI/CD, Feature Engineering, Model Deployment, MLOps

## Achievements

---

**Open Source:** Rosé Pine Theme — R/ggplot2 data visualization theme with 1.2K+ downloads on CRAN

**Technical Writing:** Medium — 13 published articles on Python, NumPy, Pandas, and data visualization

**Book (In Progress):** *Data Science Mastery: From Fundamentals to Professional Practice*

## Certifications

---

**Machine Learning**

**Applied Machine Learning in Python**

**Natural Language Processing in TensorFlow**

**SQL for Data Science**

Stanford University (Coursera)

University of Michigan (Coursera)

DeepLearning.AI (Coursera)

UC Davis (Coursera)

## Additional Information

---

- **Authorization:** Authorized to work for any US employer without sponsorship
- **Relocation:** Willing to relocate anywhere in the United States
- **Languages:** English (Professional), Bengali (Native)