SHLOKI JHA

Ann Arbor, MI · shlokij@umich.edu · (734) 489 4664 · linkedin.com/in/shloki-jha-7993591a1

SUMMARY

Aspiring data science professional with a strong foundation in analytics, programming, and statistical modelling. Pursuing a MS in Data Science at the University of Michigan, I bring hands-on experience in machine learning, predictive modelling, and data visualisation, coupled with a demonstrated ability to communicate insights and drive solutions.

EDUCATION

University of Michigan MS Data Science — Ann Arbor, MI
Relevant Coursework: Data Structures & Algorithm, Applied Machine Learning, Data Manipulation & Analysis, Statistical Inference, Information Retrieval (LLMs), Database Management Systems
Aug ’23 - May’ 25

Indian Institute of Technology BTech Civil Engineering — Roorkee, India
Relevant Coursework: Data Mining, Math for Data Science, Linear Algebra
Jul ’19 - Jul’ 23

SKILLS

Programming Languages	Python, C++, MySQL, PostgreSQL, R, MATLAB
Database Technologies	Java, MongoDB
Data Visualisation	Power BI, Tableau, Matplotlib, Seaborn
Other Software/Tools	AWS Athena, AWS Sagemaker, TensorFlow, NLTK
Co-curriculars	DOW Sustainability Fellowship (2024), Girl Up Now

RELATED EXPERIENCE

Rocket Mortgage — May ’24 - Aug ’24
Data Science Intern

• Developed predictive models for dial prioritisation of specialised leads, where the top deciles accurately captured most of the closings, improving baseline model performance and enhancing lead conversion.
• Built market share forecast models employing various statistical analysis techniques and experimented with different models like XGBoost, Decision Tree, SVM and Stacked tree regressors, contributing to more accurate business forecasts.
• Leveraged AWS Athena to efficiently extract, prepare, and build training data from scratch, streamlining the model development.
• Designed a sliding window dashboard in Power BI, delivering real-time insights to support business decisions.
• Presented key model findings to stakeholders, translating insights into actionable recommendations that drove strategic decisions.

University of Michigan - School of Information — Oct ’23 - May ‘24
Curriculum Development and Revision Aide (CDRA)

• Beta tested and helped enhance the Business SQL and Applied NLP courses in the Masters of Applied Data Science program.
• Worked with instructors and curriculum developers to implement course revisions, enhancing the student learning experience.

MedTourEasy — Sep ’22 - Oct ‘22
Data Analyst Intern

• Conducted data analysis using Excel and Python on 748 samples, visualising donor patterns and extracting actionable insights.
• Optimised predictive modelling using TPOT, improving AUC from 76% to 78.5% and streamlining model pipelines.
• Predicted blood donation trends through advanced feature engineering and TPOT-optimised models, uncovering patterns.

PROJECTS

Lyrics-Based Search Engine *Software Development, Semantic Analysis, Deep Learning, Hugging Face*
Developed a lyrics-based search engine combining statistical rankers (BM25, TF-IDF) and semantic models (Siamese BERT, LSA) to retrieve songs from partial lyrics. Evaluated on 50,000 song lyrics and 1,000 query-document pairs, Siamese BERT achieved the best performance (MAP: 0.4972, NDCG: 0.0704), improving MAP by 16.5% over the best baseline (TF-IDF: 0.4268).

Net Gains: Kicking Around Insights In Soccer Data *Data Manipulation, Statistical Analysis, Modelling*
Led analysis of 50,000 global soccer matches to identify key factors influencing team performance. Built predictive models with up to 89.5% accuracy for away scores and 83.2% for home scores using Random Forest. Created visualisations to assess team strengths, analyse league competitiveness, and provide strategic insights for analysts.

News Category Prediction *Machine Learning, Natural Language Processing, Neural Networks*
Implemented machine learning models for news category prediction using the MIND corpus (over 100k articles across 16 categories). Achieved 73% accuracy with Logistic Regression (TF-IDF), 72.4% with Neural Networks, and 72.5% with SVM. Built a sentiment analysis module using VADER and a content-based recommendation system with cosine similarity for personalisation.

Predicting Travel Mode Choice using ML *Machine Learning, Python, Jupyter Notebook*
Applied machine learning to analyse travel mode choice using the NHTS 2017 dataset (130k+ records), achieving 96.5% accuracy with Random Forest. Evaluated factors like income, fuel prices, and vehicle usage to support sustainable transportation planning.