Camellia Debnath

Data Scientist II @ Amazon

Seattle, WA 98121
+1 (857) 930-3409
camelliadebnath@gmail.com
dn-cam.github.io
linkedin.com/in/camellia-debnath

8+ years of experience in analysing large datasets, driving insights, predictive modelling, natural language processing, and software development; looking for Senior Data Scientist opportunity.

WORK EXPERIENCE

Amazon, Seattle, Oct'20 – Present

Amazon Advertising – Advertising Trust

Lead a team of Scientists, Business Intelligence Engineers, Data Engineers, and Software Developers to deliver end-to-end Data Science solutions for detecting policy violating ads using customer feedback and product attributes.

- Developed automated ticket generation framework to detect policy violating ads by building a text classification model for ad feedback, rescuing 3.5 billion yearly ad impressions
- Built an Ad feedback prioritization framework using LSTM text classification model with attention, achieving a 94% classification accuracy. This reduced the weekly average Andon (egregious advertisement) detection time by 7.5 hours.
- Used Amazon Bedrock LLMs to enable feedback tagging and product tagging to surface policy violating content, create review SOPs for emerging trends in advertising, competitor driven reviews etc.
- Built Continuous training pipeline for ML models.

Campaign and Creative Services

Lead a team of Campaign Managers and Business Analysts to deliver a capacity forecasting framework to optimize a variable cost structure.

- Designed a time-series forecasting engine using Prophet algorithm for automating capacity planning for campaign management with a month over month MAPE of 5%, this reduced average spending by 7% due to hiring overheads.

Fidelity, Boston, Jul'19 – Dec'19

Market Simulation using Deep Reinforcement Learning Agents

- Designed a framework for training Deep Reinforcement Learning (DRL) network in agent-based price-order-book simulations, market acts as the environment while a DRL network acts as an agent that places order to the market.
- Constructed DRL agent as an A2C (Advantage-Actor-Critic Network) that uses LSTM and CNN blocks for feature extraction

Citi, Pune, India, Jul'15 – Jul'18

Natural Language Processing Framework for Automated Generation of Test Cases

- Designed a framework using **Apache Jena** and **Stanford NLP** libraries for parsing plaintext, performed POS-Tagging.
- Traversed **RDF** knowledge graphs to generate **gherkin** scripts, fed into Cucumber Test runner to generate JUnit test cases
- This reduced communication latency, and increased average burndown efficiency factor by 35%

Real-Time Processing of Big Data to process Audit Data

- Designed a **Java** framework using **Apache Storm** topologies and **Apache Kafka** topics for processing trade exceptions.
- Saved processed errors into **MongoDB** for reference, and possible resolution.
- Transformed T+1 day batch processing to real-time error resolution, garnering customer satisfaction

RESEARCH FELLOWSHIP

Northeastern University, Boston
Sept'19 – May'20

Automated Problem Solving Using Neural Monte Carlo Tree Search (Research Fellow)

- Used Google DeepMind's Alpha Zero algorithm for solving combinatorial optimization problems
- Formulated game switch statistics that resulted into an important discovery about the effect of TensorFlow version difference.
- Generated game-trees to explore how an asymmetric winning strategy evolves for both proponent and opponent in the game.
- Created HSR (Highest Safe Rung) environment in OpenAI Gym for evaluation of Neural MCTS algorithm.

EDUCATION

Northeastern University, Boston, USA
Sep'18 – Aug'20

Khoury College of Computer and Information Sciences
Master of Science in Data Science

- **GPA**: 3.94/4.00
- **Relevant Courses**: Algorithms (**Teaching Assistant** for Spring'19,'20), Data Management and Processing, Supervised Machine Learning, Unsupervised Machine Learning, Foundations of Artificial Intelligence, Information Visualization.

SKILLS

Languages

Python, R, Java, C, PL/SQL

Concepts

Regression Modeling, Classification, Clustering, Language Modelling, Data Wrangling, Exploratory Data Analysis, Model Selection & Assessment, Topic Modelling, Computer Vision, Agile, Unix/Linux, Relational Databases, Natural Language Processing

Tools

Eclipse, UDeploy, RStudio, Jupyter Notebook, Autosys, Git

Libraries

numpy, pandas, scikit-learn, seaborn, TensorFlow, NLTK, tidyverse, dplyr, ggplot2, OpenCV