Camellia Debnath

SDE-II @ Amazon

(857)930-3409 • debnathc1@gmail.com • www.linkedin.com/in/camellia-debnath • github.com/dn-cam • dn-cam.github.io

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.

TECHNICAL KNOWLEDGE

Languages: Python, R, Java, C, C++, PL/SQL, Angular 2, Javascript

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

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

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

WORK EXPERIENCE

Software Development Engineer II, AMAZON, Seattle *Oct'20-Present*

Identity Services - Authentication Service

- Created restricted API management for limiting client access to sensitive customer information
- Designed mass Alexa device authentication process for hospitals and other large organizations

Intern, 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

Software Developer, CITICORP, 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 +1 day batch processing to real-time error resolution, garnering customer satisfaction

RESEARCH FELLOWSHIP

Northeastern University *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, e.g. the Highest Safe Rung.
- 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.

PROJECTS

Northeastern University *Sept'18-June'20*

Ahoybot: A Seq2Seq (Variational Auto-Encoder) Chatbot that Speaks Pirate

- Trained Seq2Seq model on the Cornell Movie Dialog Corpus converted into Pirate Language to give the chatbot a personality.
- Used LSTM-RNNs (Long Short-Term Memory Recurrent Neural Network) as the building blocks for encoder and decoder.
- Used Luong Attention Mechanism to capture the context of a conversation effectively.

Vocational Course Recommendation Engine

- Built course recommendation engine using Indeed Job dataset and course list scraped from Northeastern University Catalog.
- Used **K-Means clustering** to classify input skills to a group of jobs, extracted top keywords from the corresponding cluster.
- Calculated Pairwise Cosine Similarity of these keywords with the course data, returning the top 10 best matched courses.

Bankruptcy Prediction Using Various Classifiers

- Analyzed 5 years of Polish Companies Dataset, containing 64 econometric ratios and bankruptcy labels.
- Handled missing data using **Mean value imputation**, and **SMOTE** (Synthetic Minority Oversampling Technique) on training data for handling imbalance between two classes (bankrupt and non-bankrupt).
- Calculated **Correlation matrix**, performed **Cross-validation** for feature sub-setting and hyperparameter tuning.
- Fit Logistic Regression, Naïve Bayes, **LDA** (Linear Discriminant Analysis), **QDA** (Quadratic Discriminant Analysis), **SVM** (Support Vector Machine) and feed forward Neural Networks to compare various classification techniques.