Amanda Dsouza

dsouza.amanda@gmail.com

https://github.com/amy12xx
https://www.linkedin.com/in/amanda-dsouza-a2425137/

Profile Summary

Graduate student at Georgia Institute of Technology, in Computer Science. In my previous role, I led the text practice at Fractal. My primary areas of focus are in NLP, ML, and Information Retrieval. I applied these techniques to build scalable solutions for industry. My research interests are in Reinforcement Learning and NLP.

In my previous positions, I have managed a number of IT (C#, Java) projects on site for Fortune 500 companies in the US and Europe.

I contribute frequently to open source projects (scikit-learn, dirty_cat) and am an active member of data science communities (Data Umbrella, Machine Learning Tokyo, RL Discord). I have mentored at Data Umbrella and Humble Data sprints.

Technologies & Education

Python, Pytorch, OpenAI Gym, Scikit-Learn, Numpy, SciPy, Pandas, NLTK, Spacy, Flask, Tornado. C#, SQL.

Current student, MS Comp. Science, Georgia Tech. Bachelor Computer Engineering, PCCE, Goa, India.

Experience

Lead Data Scientist

Fractal.ai July 2016 – Dec 2020

I lead teams in Text Mining / NLP and ML projects. I also support new and existing client proposals, and present Fractal's text capabilities to C-level executives of various clients.
I am responsible for the research and development of Fractal's Text Analytics IP solution (which currently comprises of modules for Topic Modeling, Sentiment Analysis, Entity Resolution and Context Extraction) and Fractal's text practice, in areas of Conversational Agents, Text Summarization, Question-Answering and Cross-Lingual pipelines.

Key project / client engagements

- Developed a news events trigger engine for a large financial institution to enable its relationship managers (RM's) to better understand their customers. The engine summarized news articles relevant for an event category, and related to specific companies the RM's are interested in. Key components of the solution included entity resolution, named entity recognition, text classification and ranking.

- Designed and implemented a multi-label classification system for automating the tagging of customer's calls in a customer service center, for a large retail client. Using historical voice (audio recordings) data and text (emails, chat), built machine learning models to predict tags in Dutch language. Key components included data anonymization using NER, STT using google API's, and building text classification models in Dutch language.

- Designed and implemented an end-to-end Information Retrieval system for a health & wellness company, for providing the backbone for search, question-answering systems. The engine recommended products based on user queries, including long form questions, and provided supporting wellness/health information for the question. Key components included harmonizing data from different sources, preprocessing query and documents for maximizing retrieval, fast K-documents retrieval using LSH, and ranking.

- Built an aspect-based sentiment analysis solution for customer reviews of consumer products company, in English and French, at the level of brand, and review category (price, quality etc.) Built aspect-based models using both unsupervised (using dependency parsing rules) and supervised (classification models) methods.

- Built an online adaptive histogram model for detecting anomalies in large scale printer sensor data on multivariate data.

- Built a trend trigger system on corporate financial (time-series) data, and devised methods to automatically detect seasonality in data, as well as predict trend reversals and momentum changes, using techniques of change point detection, and technical analysis indicators.

- Presented Fractal's capabilities on real time text analytics at Microsoft's vendor conference in Redmond WA, and on Text mining case studies at Data Hack Summit 2017, Bangalore. Conducted a workshop for a large retail client on Artificial Intelligence.

- Worked on RL models, developed DQN model for training a lunar lander agent, and implemented other algorithms (Q-Learning, Correlated Q, Policy Gradients). Built a RL agent using MuZero for text based game Zork.

Technology Lead

Infosys Technologies May 2005 – May 2016

Data scientist for Infosys's Machine Learning Center of Excellence, responsible for implementing data mining and statistical machine learning solutions to various business problems. My previous roles were Technology Lead and Software Developer.

- Built classification models for detecting driver pattern based on only GPS route data, for usage in insurance model. Applied RDP (Ramer-Douglas-Peucker) for route approximation and Kalman filters for reducing noise on GPS data to build smoother feature set.

- Conducted predictive analytics of client's site planning data for master data automation. Predicted output parameters from large scale & high dimensional data, using ensemble classifiers and regressors. Achieved over **95% accuracy** for most output parameters. Built a model for detecting churn for telecom client data, with a 65% accuracy of predicting churn. Built models for predicting loan default with over **96% accuracy**.

- Created a text classification model for determining text categories from social media text (Twitter, Facebook…), for Voice of Customer application.

- Provided architecture solution for several key projects and RFPs. I have managed several IT projects over the years, in Agile, Waterfall and Iterative methodologies. I was the on-site coordinator for a major Aero OEM project, with a 15-member offshore development team. I was the business analyst and scrum master for a 10-member development team for a major Auto manufacturer.

Courses Completed (Georgia Tech):

Knowledge Based Artificial Intelligence
Machine Learning for Trading
AI for Robotics
Reinforcement Learning
Artificial Intelligence
Simulation
Deep Learning
Deterministic Optimization