Amanda Dsouza

206-679-0825 | dsouza.amanda@gmail.com | https://www.linkedin.com/in/amanda-dsouza-a2425137/

Profile Summary

Lead Data Scientist, with 6 years of experience in NLP/ML. I led and mentored a team of DS and Senior DS' at Fractal.ai, working on natural language processing and text mining problems, for industry, and applied research. I was hands-on, on the projects I led.

I am currently an independent researcher, and recent graduate from Georgia Tech, in Computer Science My research interests lie broadly in Reinforcement Learning, Natural Language Processing and Representation Learning. My current research projects are on Imitation Learning, Question Generation and Generalization. I am also motivated in finding interesting ideas from ML adjacent fields. I write on these topics at https://amy12xx.github.io/.

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 have contributed to open source projects (scikit-learn, dirty_cat) and am a member of data science/research communities (Data Umbrella, ML Collective, Cohere 4 AI, RL Discord). I have mentored at PyData, Data Umbrella and Humble Data sprints. I just started co-organizing the RL reading group at Cohere 4 AI.

Technologies & Education

Python, Pytorch, Scikit-Learn, Numpy, Pandas, OpenAI Gym, IsaacGymEnvs, Stable Baselines3, Weights & Biases, Tensorboard, Hugging Face, Git.

Master of Science in Computer Science (ML) - Georgia Institute of Technology, Atlanta, GA
Bachelor of Engineering (Computer) - PCCE, Goa, India

Experience

Independent Researcher, Graduate studies

Jan 2021 - Present

- Developing an Imitation Learning approach, extending Generative Adversarial Imitation Learning, with Vector Quantization for improved representation learning. Working with a PhD (Robotics) candidate. Testing on IsaacGym environments.

- Evaluating the effectiveness of different prompt-based methods for generation of deep cognitive questions using Large Language Models, and testing out a similarity based technique for auto-generating deep questions.
- Surveying generalization experiment design in RL, and related fields.

Lead Data Scientist

Fractal.ai   July 2016 – Dec 2020

I lead teams in Text Mining / NLP and ML projects for large companies in Retail, Financial Services, FMCG. I also supported new and existing client proposals, and presented Fractal's text capabilities to C-level executives of various clients. I was responsible for the research and development of Fractal's Text Analytics IP solution, and led applied research sessions in the NLP practice.

Key project / client engagements

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

- 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 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 a 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 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.

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 patterns based on only GPS route data, for usage in insurance models.

- 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 for Voice of Customer application.

- Provided architecture solutions 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 Boeing, with a 15-member offshore development team. I was the business analyst and scrum master for a 10-member development team for Toyota Motor Europe.