Montreal,QC - Canada
+1 (780) 200-8010
markosviggiato@gmail.com
markosviggiato.github.io/
linkedin.com/in/markosviggiato

Markos Viggiato

Highlights of Qualifications

- Experience with applied research in industry using large-scale data
- Demonstrated ability to critically identify, prioritize and solve problems
- Proven record of collaboration with different research teams and stakeholders (papers [P1, P3]), paper writing, and communication skills (demonstrated by conference paper presentations [P1, P4])
- Technical expertise: NLP (language modeling, text classification, named entity recognition, sentiment analysis), statistical modelling, prediction and explanatory models, unstructured data analysis, cloud platforms (GCP, Snowflake), SQL, prototyping skills

Education

Jan-2019–Present **PhD in Electrical and Computer Engineering**, *University of Alberta*, Canada
- Applied research in Natural Language Processing. GPA: 4.0 (out of 4.0)

Mar-2017–Dec-2018 **MSc in Computer Science**, *Federal University of Minas Gerais*, Brazil
- Machine Learning and Data Mining for Software Engineering. GPA: 9.0 (out of 10.0)

Mar-2011–Dec-2016 **Bachelor in Control and Automation Engineering**, *Federal University of Minas Gerais*, Belo Horizonte, Brazil. GPA: 7.6 (out of 10.0)

Industry Experience

Oct 2020–Present **NLP Research Intern**, *Prodigy Education*

I collaborate with data scientists, product managers, and engineers to build an end-to-end pipeline that uses ML/NLP techniques together with large-scale data to automatically extract knowledge from existing testing scenarios in natural language, improve new testing scenarios, and support user-centric, data-driven test execution **(Python, PyTorch, Transformers, Snowflake, Prototyping)**

- Developed a deep learning-based system using an ensemble of transformer-based, zero-shot models and embedding techniques to automatically label 7000+ game testing scenarios in natural language with an accuracy of 76%.
- Combined statistical (n-grams) and neural (fine-tuned BERT) language models to build an application to accelerate the design of high-quality game testing scenarios in natural language by suggesting better words in real-time during testing design. The application is deployed as a browser extension and collects usage data to continuously monitor and improve the models.
- Built a deep learning-based approach using Word2Vec and pre-trained BERT and Sentence-BERT models to support the identification of similarities and redundancy in 7000+ testing scenarios with an accuracy of 86%.
- Built a dashboard to analyze the usage of 100+ game features and automatically retrieve related testing scenarios in natural language, which provides in-depth insights to understand the behavior of 10+ millions of users, marketing effectiveness, user retention and conversion, and high-priority areas for testing.

Research Experience

Jan-2019–Present **PhD Researcher**, *University of Alberta*

Applied research in machine learning and NLP **(Python, PyTorch, Transformers, Java)**

- Investigated ensembles of zero-shot and few-shot techniques to automatically label testing documents in natural language.
- Investigated trade-off between statistical and neural language models using different evaluation metrics (*Perplexity* and *Accuracy@k*) to build a system that automatically analyzes testing scenarios in natural language and provides improvement recommendations.
- Implemented a sentiment classification pipeline to analyze 12M of game reviews using GPT-3 and ready-to-use tools.
- Modelled user behavior to build explainable win prediction models (XGBoost, Random Forest, Logistic Regression) for the Dota 2 game using the SHAP interpretability technique and achieved a performance of 86%.

Jan-2017–Dec-2018 **MSc Researcher**, *Federal University of Minas Gerais*

Applied research in machine learning and data mining for software engineering **(Python, R, Java)**

- Modelled the behaviour of developers using statistical and explanatory models and leveraged frequent itemset algorithms to identify co-evolution of changes in software development.
- Collaborated on a project to build explainable prediction models for software defects using XGBoost and SHAP values and improved the prediction accuracy by 15%.

Selected Publications

P1 **Using Natural Language Processing Techniques to Improve Manual Test Case Descriptions.** Markos Viggiato, Dale Paas, Chris Buzon, Cor-Paul Bezemer. *The 44th IEEE/ACM Conference on Software Engineering (ICSE 2022)*

P2 **Identifying Similar Test Cases That Are Specified in Natural Language.** Markos Viggiato, Dale Paas, Chris Buzon, Cor-Paul Bezemer. *IEEE Transactions on Software Engineering*, 2022

P3 **What Causes Wrong Sentiment Classifications of Game Reviews?** Markos Viggiato, Dayi Lin, Abram Hindle, Cor-Paul Bezemer. *IEEE Transactions on Games*, 2021

P4 **Trouncing in Dota 2: An Investigation of Blowout Matches.** Markos Viggiato, Cor-Paul Bezemer. *The 16th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2020)*

Additional Information

Awards

- Alberta Innovates Graduate Student Scholarship (Jan 2020 - present). 3-year duration scholarship
- Alberta Graduate Excellence Scholarship (AGES) (Sep 2019)

Leadership positions

- Lead researcher in an industry-academia collaboration project between the University of Alberta and Prodigy Education

Other Skills

- Experience with project management, git, SQL, Snowflake, Google Cloud Platform (GCP), Jupyter Notebook, Sklearn, Spacy, PyTorch
- Experience with the following languages: Python, R, C/C++, Java