Wilson Estécio MARCÍLIO-JR
Presidente Prudente/São Paulo, Brazil
+55 18 998052957
wilson_jr@outlook.com
website | github | linkedin

SUMMARY

Wilson has Bachelors and Masters in Computer Science, and is now a Ph.D. candidate in Computer Science at the São Paulo State University – Brazil. His expertise lies in unsupervised learning and in the intersection of Information Visualization and Machine Learning, which consists of helping humans in the process of understanding machine learning algorithms.

EDUCATION

SÃO PAULO STATE UNIVERSITY - UNESP *2019-2022*

PH.D. IN COMPUTER SCIENCE

• Development of unsupervised learning algorithms focusing on the intersection of machine learning and visualization techniques. Developing novel:

- dimensionality reduction techniques;
- feature selection techniques using interpretable strategies;
- topic extraction techniques;
- visualization approaches to explain decisions of machine learning models.

SÃO PAULO STATE UNIVERSITY - UNESP *2016-2018*

MSC IN COMPUTER SCIENCE

• Focusing on the development of visualizations and methods for high-dimensional data analysis.

SÃO PAULO STATE UNIVERSITY - UNESP *2012-2016*

B.S IN COMPUTER SCIENCE

• 1st place in the regional phase of the programming contest in 2014.

EXPERIENCE

IBM *Jun/2021 to -*

APPLICATIONS DEVELOPER

SÃO PAULO STATE UNIVERSITY - UNESP *Feb/2019 to April/2020*

TECHNICAL TRAINING

• Working on the research and development of a Visual Analytics approach to explore text and image collections.

- Convolutional Neural Networks for feature extraction;
- Using Natural Language Processing and Language models for text analysis;
- Implementation of novel visualization strategies for exploratory data analysis;
- Technologies: scikit-learn, Keras, Pytorch, D3.js, Python, JavaScript, Java.

DALHOUSIE UNIVERSITY, HALIFAX/CANADA *March/2018 to April/2018*

VISITING RESEARCH STUDENT

• Working on the research and development of a Visual Analytics approach to explore image collections.

- Convolutional Neural Networks for feature extraction;
- Dimensionality reduction for visualization of feature spaces;
- Clustering and sampling strategies for similarity analysis;
- Implementation of novel visualization strategies for exploratory data analysis;
- Technologies: scikit-learn, Keras, D3.js, Python, JavaScript, Java.

DUKE ENERGY CORPORATION *April/2015 to November/2015*

MACHINE LEARNING DEVELOPER INTERN

• Working in the development of machine learning methods for detecting unauthorized changes in the surroundings of hydroelectric plants.
  - Technologies: Python, scikit-learn, and PostgreSQL.

Skills

Expert: Data Visualization | Data Analysis | C++

Experienced: Machine Learning | Scikit-learn | Python | Java

Skillful: NLP | Keras | Git

Beginner: Pytorch | Amazon SageMaker

Languages

Native: Portuguese ***Fluent:*** English

Courses

- Deep Learning from DeepLearning.AI
- TensorFlow developer from DeepLearning.AI
- AI for Medicine from DeepLearning.AI

Projects

ClusterShapley *May/2021 to Now*

https://github.com/wilsonjr/ClusterShapley

• ClusterShapley is a technique to explain non-linear dimendionality reduction results.
  - It produces the feature contributions for the clustering result after dimensionality reduction.

RADAR COVID-19 *Feb/2020 to Now*

https://covid19.fct.unesp.br/coronavirus

• RADAR COVID-19 is a project of the Universidade Estadual Paulista (UNESP) for monitoring COVID-19 in the São Paulo state. We use Visual Analytics and Data Mining techniques to understand the dissemination of COVID-19.
  - I am responsible for the following activities:
  - Natural Language Processing on tweets about COVID-19 symptoms;
  - Tweets classification using language models (BERT);
  - Development of Visual Analytics strategies for exploratory data analysis.

SHAP_FSelection *Feb/2021 to Now*

https://github.com/wilsonjr/SHAP_FSelection

• SHAP_FSelection is a feature selection technique developed based on the following characteristics:
  - Reduce the number of feature required for a machine learning model under classification or regression tasks;
  - Provide explanations about the results.

SADIRE *Feb/2021 to Now*

https://github.com/wilsonjr/SADIRE

• SADIRE is a scatter plot sampling technique developed based on the following goals:
  - Reduce the number of points in a result of dimensionality reduction;
  - Maintain relevant information for further analysis.

Selected Publications

Journals

- **Marcílio-Jr W.E.**, Eler D.M.; *Explaining dimensionality reduction results using Shapley values*, 2021. Expert Systems with Applications, Elsevier.

- **Marcílio-Jr W.E.**, Eler D.M., e Garcia R.E.; *Contrastive analysis for scatter plot-based representations of dimensionality reduction* **(under review)**, 2021. arXiv.
- **Marcílio-Jr W.E.**, Eler D.M., e Breve F.; *Model-agnostic interpretation by visualization of feature perturbations* **(under review)**, 2021. arXiv.
- **Marcílio-Jr W.E.**, Eler D.M.; *SADIRE: a context-preserving sampling technique for dimensionality reduction visualizations*, 2020. Journal of Visualization, Springer.
- **Marcílio-Jr W.E.**, Eler D.M., Garcia R.E., e Pola I.R.V.; *Evaluation of approaches proposed to avoid overlap of markers in visualizations based on multidimensional projection techniques*, 2019. Information Visualization, SAGE.

Conferences

- **Marcílio-Jr W.E.**, Eler D.M.; *From explanations to feature selection: assessing SHAP values as feature selection mechanism*, 2020. 33rd Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE.
- **Marcílio-Jr W.E.**, Eler D.M., Garcia R.E., Correia R.C.M., e Silva L.F.; *A hybrid visualization approach to perform analysis of feature spaces*, 2020. 17th International Conferenceon Information Technology-New Generations (ITNG), Springer.
- Rodrigues, R.B.M., **Marcílio-Jr W.E.**, Eler, D.; *Entropy-Based Filter Selection in CNNs Applied to Text Classification*, 2020. 9th Brazilian Conference on Intelligent Systems (BRACIS), IEEE.
- Eler D.M., Batista, M.P., Garcia R.E., Pereira, D.R., **Marcílio-Jr W.E.**; *Visual Approach to Support Analysis of Optimum-Path Forest Classifier*, 2019. 8th Brazilian Conference on Intelligent Systems (BRACIS), IEEE.
- Rodrigues, G.E., **Marcílio-Jr W.E.**, Eler, D.; *Data Classification: Dimensionality Reduction Using Combined and Non-combined Multidimensional Projection Techniques*, 2018. 7th Brazilian Conference on Intelligent Systems (BRACIS), IEEE.
- **Marcílio-Jr W.E.**, Eler D.M., Garcia R.E.; *An approach to perform local analysis on multidimensional projection*, 2017. 30th Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE.