Matheus Agostini Ferraciolli

An Agricultural Engineer turned data scientist by crop yield modeling. I started using machine learning to study the effects of spatial autocorrelation in empirical modeling of sugarcane yield, coffee disease occurrence prediction and soil classification models. Outside of agriculture, I have developed projects on fraud detection, lawsuit outcome prediction, customer churning, logistics transportation, car GPS tracking, students performance on online education and survey data analysis.

Education

2018-2020 (expected) **MSc, Data Science**; University of Campinas (Unicamp)

Object detection on greenhouses for crop monitoring and yield estimation

2012-2017 **BSc, Agricultural Engineering**; University of Campinas (Unicamp)

Experience

2017-Now **Work**: Exemplaria Solutions

Data Scientist. Working on machine learning application in different fields. From ETL routines to model deployment.

2016-2017 **Research**: Bolsista de Iniciação Científica at CNPq

Effects of spatial autocorrelation in crop yield modeling. Not accounting for spatial autocorrelation between observations, models severely underestimated the error in new data. Using blocks of observations based on location helps better estimate the prediction error in fields that were not used when building the models.

2016-2017 **Research**: Intern at Brazilian Agricultural Research Corporation

Alternatives for automatic soil classification. Research on soil type classification using chemical and physical properties using decision trees.

Publications

2018 **Neglecting spatial autocorrelation causes underestimation of the error of sugarcane yield models**

https://doi.org/10.1016/j.compag.2018.09.003a

We implemented a spatially-aware protocol and compared it with the naive approach of assuming independence between samples. The protocols were applied through all the model development pipeline: data splitting for hold-out sets, feature selection, cross-validation for model adjustment and model evaluation. Three different machine learning techniques were used to create models in each

protocol. The resulting models were evaluated both in the validation set created by each protocol and in a manually created independent set. This independent set ensured there was no auto-correlation between the samples used for modeling. We showed that assuming independence when modeling yield leads to underestimating model errors and overfit during model adjustment.

Skills

- **Applied Machine Learning:** Experience and understanding of the necessary steps to test and implement different machine learning techniques for feature engineering, feature selection, cluster, classification and regression tasks. Also on proper feature engineering, model training and cross-validation.
- **Programming Languages:**
- **Python - Advanced:** Experience with numpy, scipy, pandas, scikit-learn, plotly and otherlibraries used for data manipulation, visualization and algorithms for different learning tasks.
- **R - Advanced:** Experience with tidyverse packages and mlr for large dataset reading and manipulation, visualization and different machine larning algorithms implementations and can write reports in Rmarkdown.
- **Bash:** Can automate tasks for file editing and pipelines using makefiles.
- **Deep Learning**:
  - Completed **certified course** Neural Networks and Deep Learning by deeplearning.ai on Coursera.
  - Completed **certified course** Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization by deeplearning.ai on Coursera.
  - Completed **certified course** Structuring Machine Learning Projects by deeplearning.ai on Coursera.
- **Cloud computing:** Experience setting up and using virtual machines and some experience using Azure MLstudio.
- Also have experience with Linux, IPyhton/Jupyter Notebooks, Markdown, LaTex, Kaggle Kernels and competitions.
- **Languages**:
  - Portuguese (native speaker)
  - English (fluent)

---

matheusferraciolli@gmail.com • +55 (19)9 9795 6737 • 27 years old
Rua Angelo Barijan - Sumaré, SP - Brazil.