Neoway Business Solutions
There is more to know
R. Patrício Farias 131, Florianópolis, Brazil
+55 48 3333-2030

Thesis Work - Automated Feature Engineering

Background

Neoway is the largest Big Data and Analytics company of Brazil, headquartered in Florianópolis and with operations in Latin America, USA, and Europe. We capture and integrate data from hundreds of offline (form-like) and online (internet event-based like) sources into our private data framework. In Brazil, we hold the most consistent dataset of all companies in the country, covering a broad scope of information: from firmographics to debt, assets, legal processes, construction work, trade, shareholder network, human resources, fleet, intellectual property, environmental licenses, e-commerce usage, etc. With this information, we help customers in any segment optimize decisions through data in our Multi-Market Intelligence System.

Our Data Analytics team makes use of these data to extract actionable insights through Data Science and Machine Learning. We conceive, design and implement models which are served as SaaS and MLaaS in our platform. Our models set Neoway apart from competitors, bringing us an edge in business and technology. While the benefits of machine learning are undeniable, its broad adoption and impact are limited due to the long development lifecycles involved. We believe and invest in the power of automation and would like to further explore this with Machine Learning.

Project Description

We identified that Feature Engineering, the task of creating variables with predictive power to serve as input to Machine Learning, can take up to 80% from the development cycle of our models. At the same time, we noticed that many features could find use across projects but there was no well-defined way to share, govern and monitor them. For that reason, we created a Features Framework, which provides a consistent environment for our engineers to create, share and govern our features. This has considerably improved our ability to quickly create and deploy models based on existing features.

The generation of new features is still a rather manual process. As the next step in our Features Framework, we would like to do this in a semi or fully automated way. Feature creation is widely based on a few types of functions: transformations (applied on a single table, e.g. percentile, rank), aggregations (those applied across tables, e.g. mean, count), and those applied through time (e.g. changes, increments, decrements). The problem of generating every possible feature is of combinatorial nature, depending on the pool of transformations, aggregations and the relationship between the data entities. Since 2015 some efforts have been made into tackling this problem through automation, and a number of tools have been made available such as FeatureTools. We have experimented with these tools, and despite initial success, they have not achieved our requirements on scalability and integration with our existing Features Framework.

About Florianópolis

Florianópolis is a subtropical island south of Brazil. Known for its high quality of life and natural beauty, it is also the innovation capital of Brazil. Hosting over 20% of all startups in Brazil, it has been named "The Silicon Island" of Brazil. You can be sure to find a strong entrepreneurial and forward-thinking community in a laid back atmosphere.

Neoway Business Solutions
There is more to know
R. Patrício Farias 131, Florianópolis, Brazil
+55 48 3333-2030

What you will do

As a Master thesis worker at Neoway, your mission will be to help us deliver Automated Feature Engineering at scale to our Data Analytics team. This entails:

- Literature survey on Automated Feature Engineering.
- An immersive experience in our data and our Features Framework.
- Building an entity-relation (ER) map for our main data entities.
- Conceptualize a scalable architecture together with our senior staff.
- Conceive and implement a tool for automated feature engineering based on the ER map and the historical data relationship.
- Conceive and implement tools to monitor and remove redundant features.
- Verify the practical value of the generated features in real-world use cases.

Who you are

- Undergraduate student in a quantitative discipline such as Computer Science, Engineering, Mathematics, Statistics.
- You have taken sensible steps towards a Data Science career and have a strong will to fully develop yourself as a Data Scientist/Software Engineer.
- You have a strong quantitative and analytical capacity, a drive for impact and enjoy working in a high-performance environment.
- You have previous coding experience (preferably in Python), understand the importance of quality code and are willing to develop your software engineering skills.

What we offer

- Compensation in the form of a scholarship.
- Round-trip tickets from/to Sweden to/from Florianópolis, Brazil.
- Mediation with a local university (UFSC) regarding visa and mobility issues.

Application

Your application should contain a motivation letter, CV, transcripts and possible references. Please send your application to André Carvalho Bittencourt, andre.bittencourt@neoway.com.br