



Data Analytics

Micro Project - Predictive Process Monitoring

Introduction:

You're a team of employees working for a large process mining company. Your company receives a contract to help the Swedish car manufacturer Volvo with their incident management process in their IT department. Your team is provided with the "Volvo.xlsx" file by Volvo and your task is to now get an insight into the problem and try to create a model to predict the future behavior of this process based on the log you were provided. It contains the event data produced by first six process activities and the outcome of the process instance. Also, you need to find out which service level (first, second, third) will tackle the problem at the end?

Tasks:

Your project should include the following points and should also be included in your project presentation:

- **Data Understanding and Cleanup:** Describe your dataset and give an insight into the data you're working with. This should also include visualization and giving an insight into its most important features.
- **Choosing a model:** Choose a model for your specific problem. This should include a definition of the prediction problem. Compare different models and explain which model you have chosen and why it's the best suited one for your project.
- **Evaluation:** Talk about the experiences you have collected during your project. What were your biggest challenges, how did you overcome them and talk about your biggest lesson learned during the project. evaluate the trained model on unseen data.
- **Creating a presentation:** Your presentation should include and represent the most important aspect of the tasks named above. It should also include a brief description of the single steps that you've taken to achieve your goals.

(Note: This project is meant to give you a head on approach on data analytics. Also, creative solutions such as feature engineering, the collection of additional data, etc. are welcome and will have a positive impact on the amount of bonus points you will earn from this project.)

Additional information:

Because of the complexity of your task, you are also provided with additional research material. This dataset is also part of the BPI-Challenge 2013, which might come in handy when you should be stuck.