Developing a KPI to the Slip-to-Slip Connection Time in a Drilling Operation

Pedro Grosman

Summary

What is delivered?

I delivered here a complete data-driven solution to monitor crew performance in drilling operations from an Oil & Gas company.

What the solution can do?

This solution provides an automated way to monitor the crew performance saving the company time, and money.

1.

Introduction

Understanding the business questions.

> Measurement of performance is vital to business success and most performance studies are related to the so-called key performance indicators (KPI)

Slip to Slip Connection Time

The slip is a device used to grip and hold the upper part of a drill pipe to the drill floor.

Main Goals

- Doing an exploratory data analysis to extract useful information for the development of the previously-mentioned KPI

- Creating a machine learning model to monitor the performance of rig crews in drilling operations.

2.

Exploratory Data Analysis

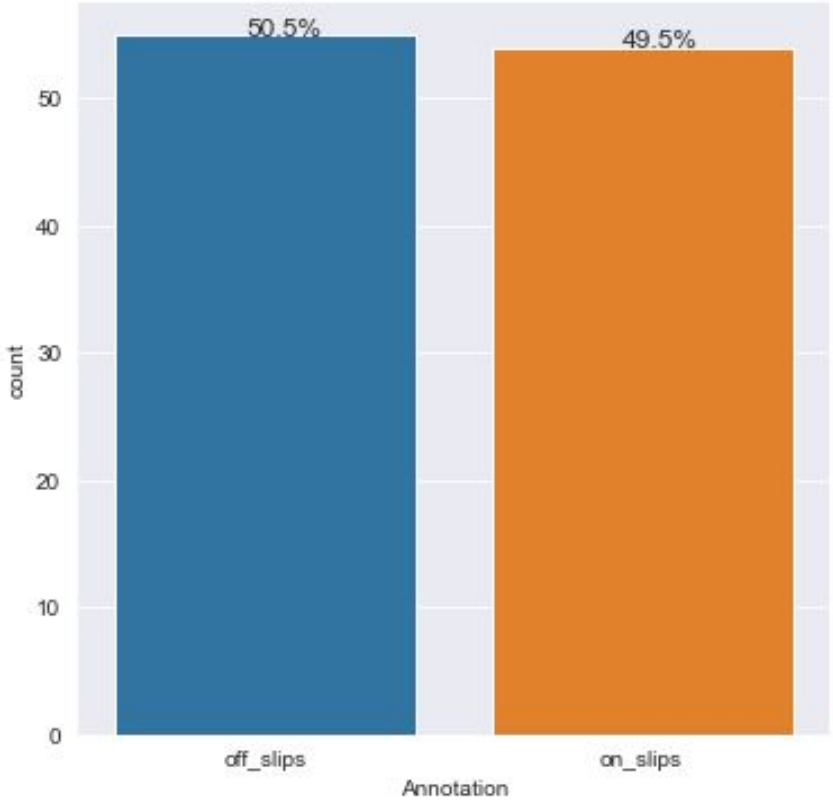
When large-sized tabular data starts to make sense

Variables

- Timestamp (ms)
- Bit depth (m)
- Fluid flow (gpm)
- Hook load (klbf)
- Block position (m)
- Rotary speed (rpm)
- Torque (klbf-ft)
- Hole depth (m)
- Weight on bit (klbf)
- Annotation (whether the slips are on or off)

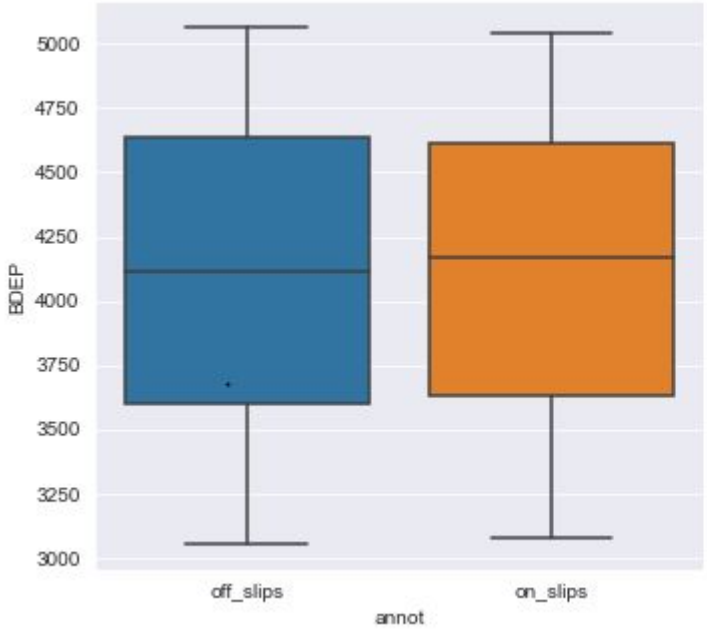
Occurrence Analysis

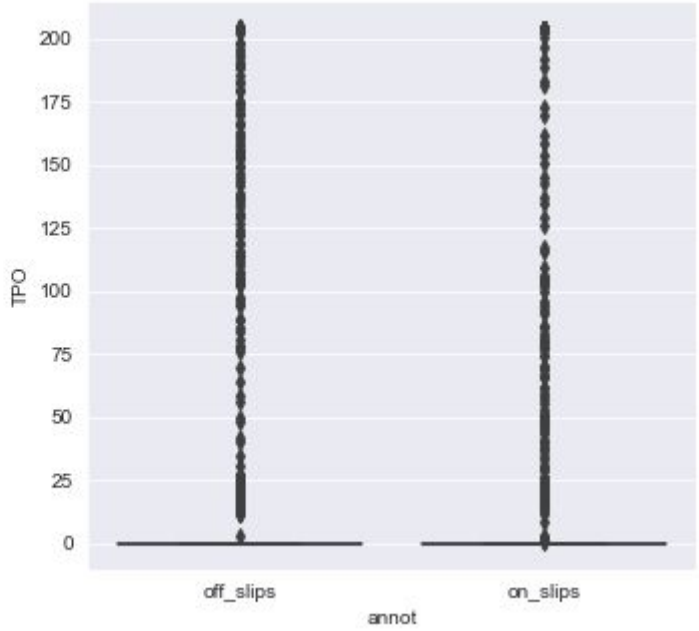
The number of '***off_slips***' annotations against the number of '***on_slips***' annotations.



Numerical Features Analysis

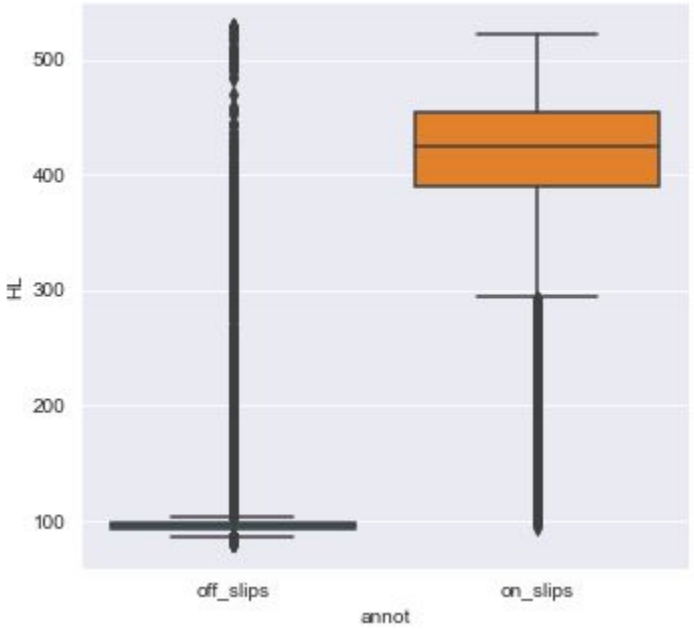
Evaluating central and dispersion metrics

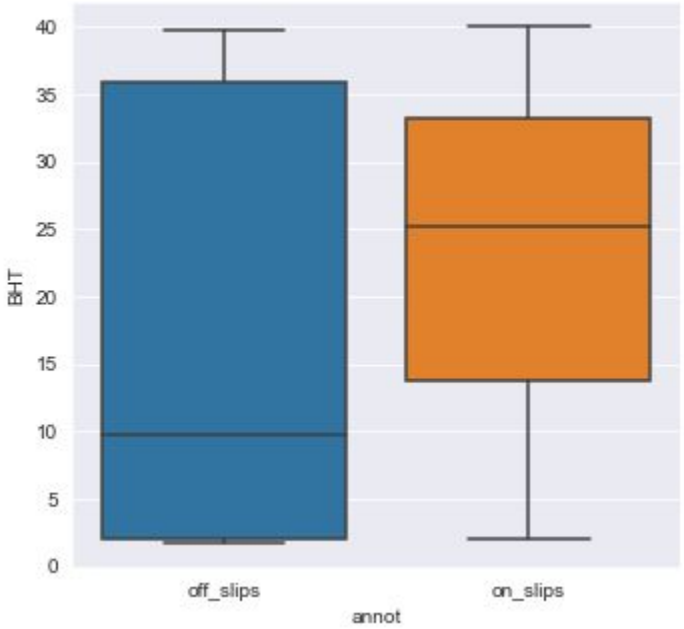




Numerical Features Analysis

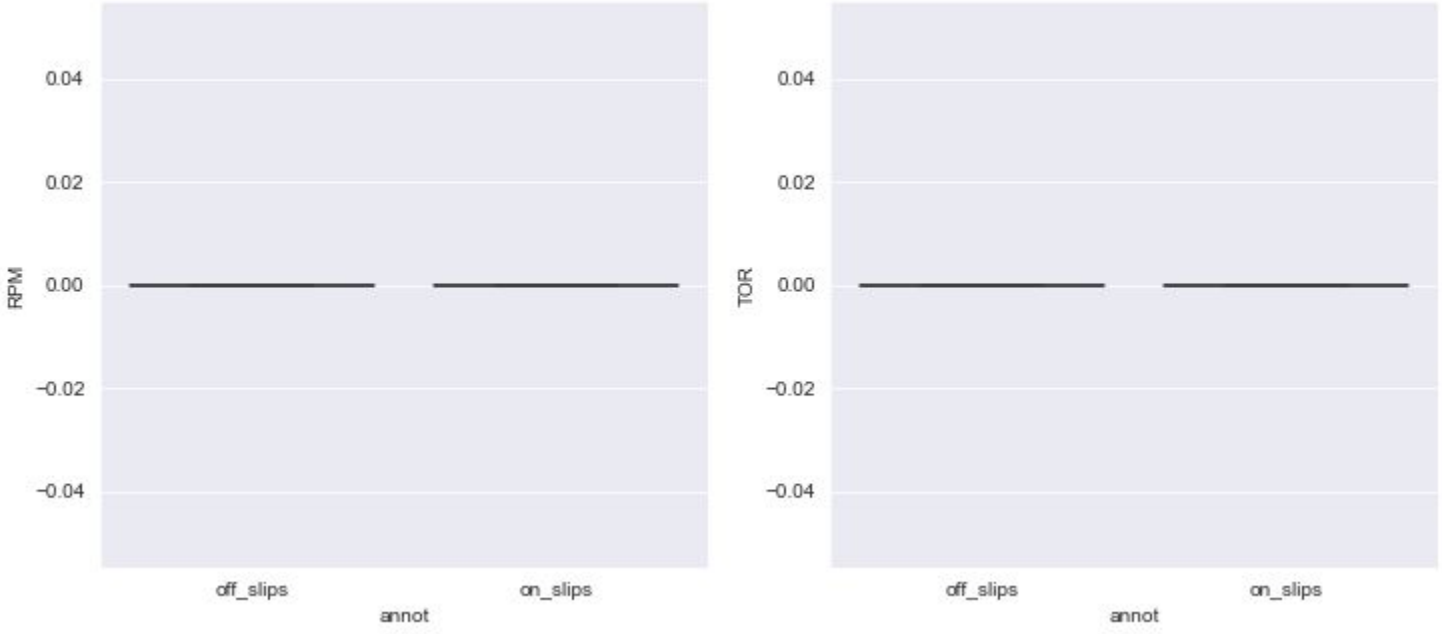
Evaluating central and dispersion metrics





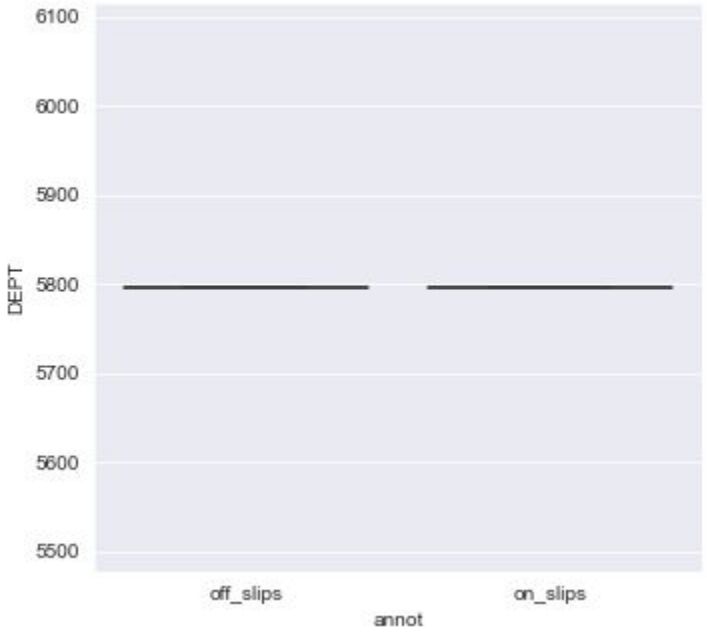
Numerical Features Analysis

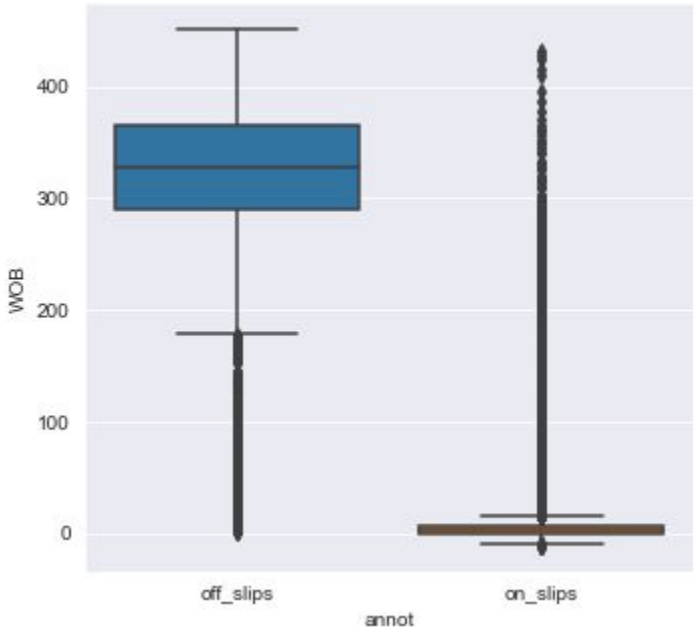
Evaluating central and dispersion metrics



Numerical Features Analysis

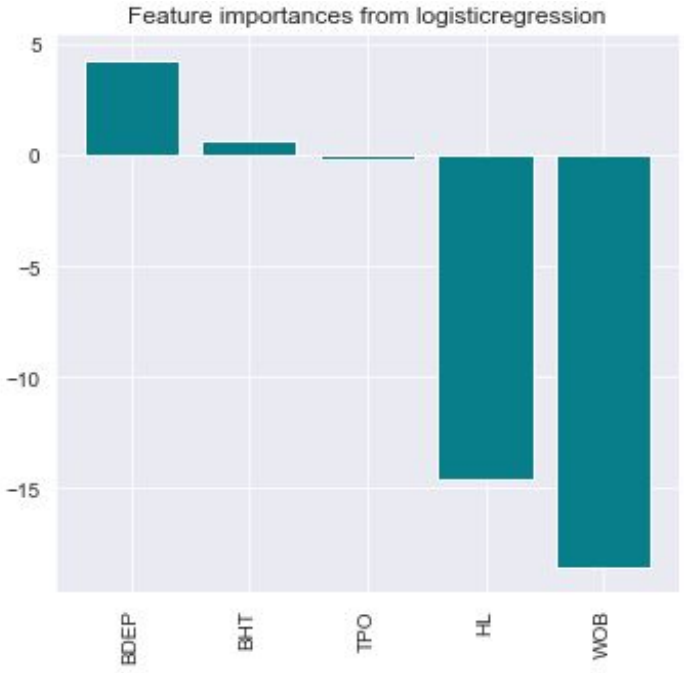
Evaluating central and dispersion metrics

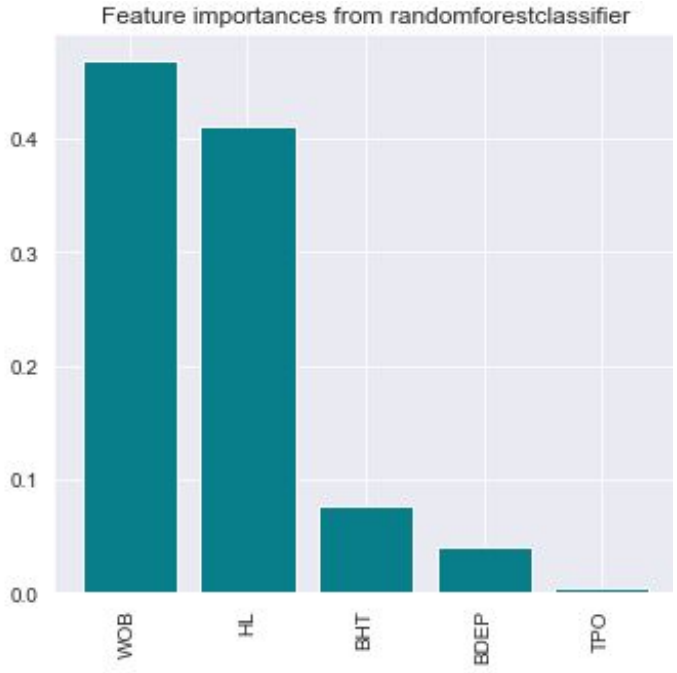




Relevance Analysis

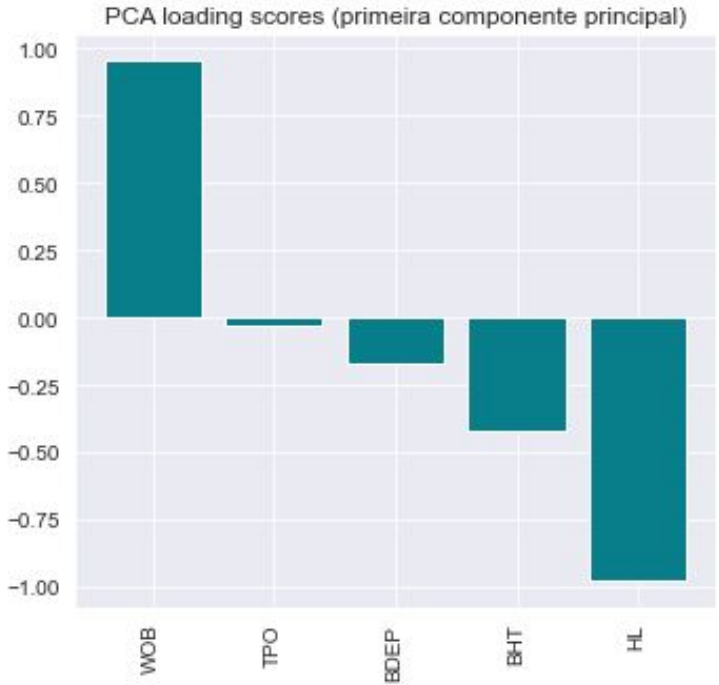
Measuring the importance of the variables using machine learning algorithms.

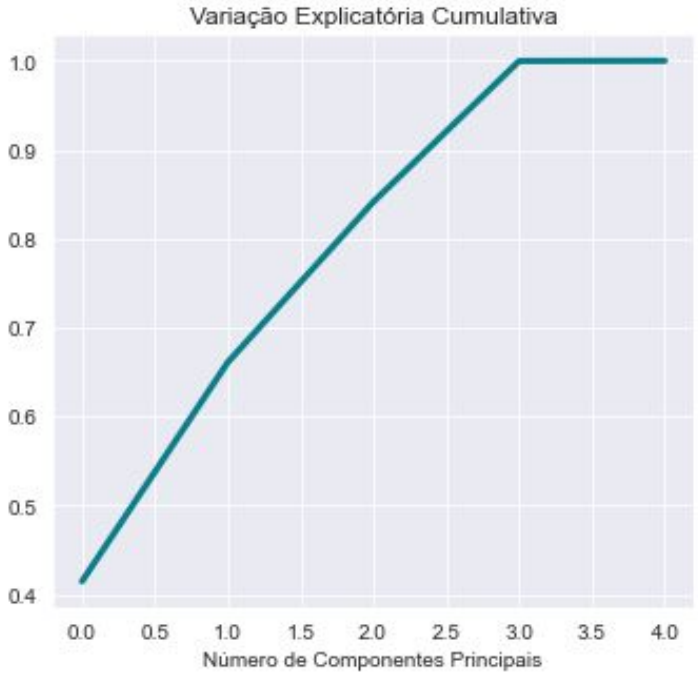




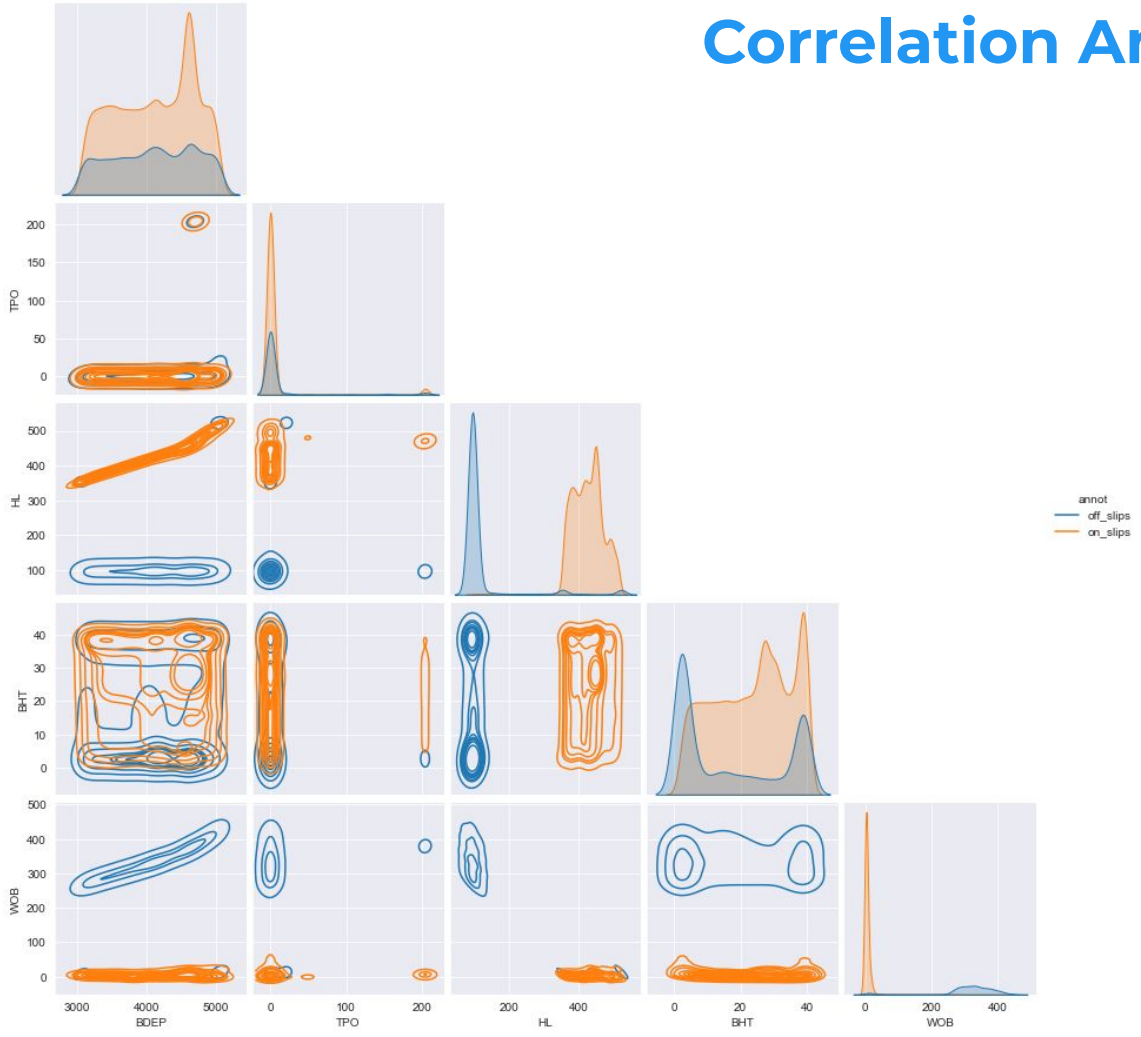
Relevance Analysis

Measuring the importance of the variables using machine learning algorithms.

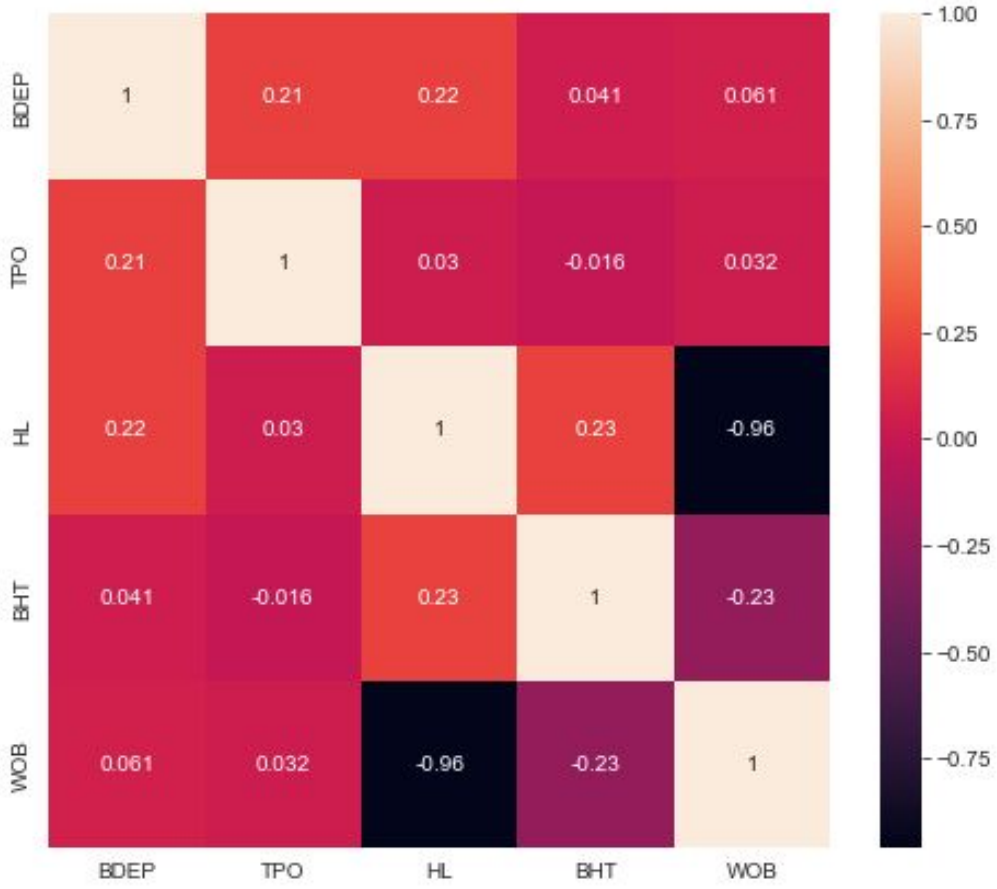




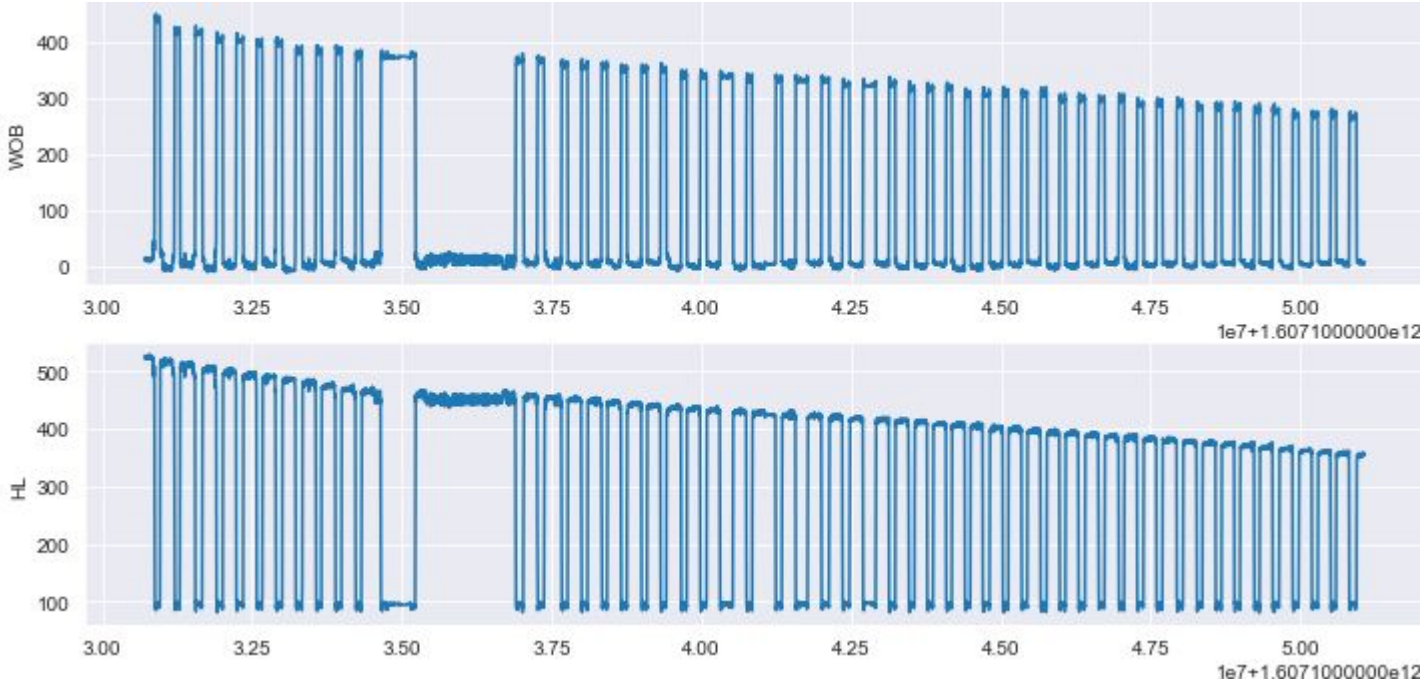
Correlation Analysis



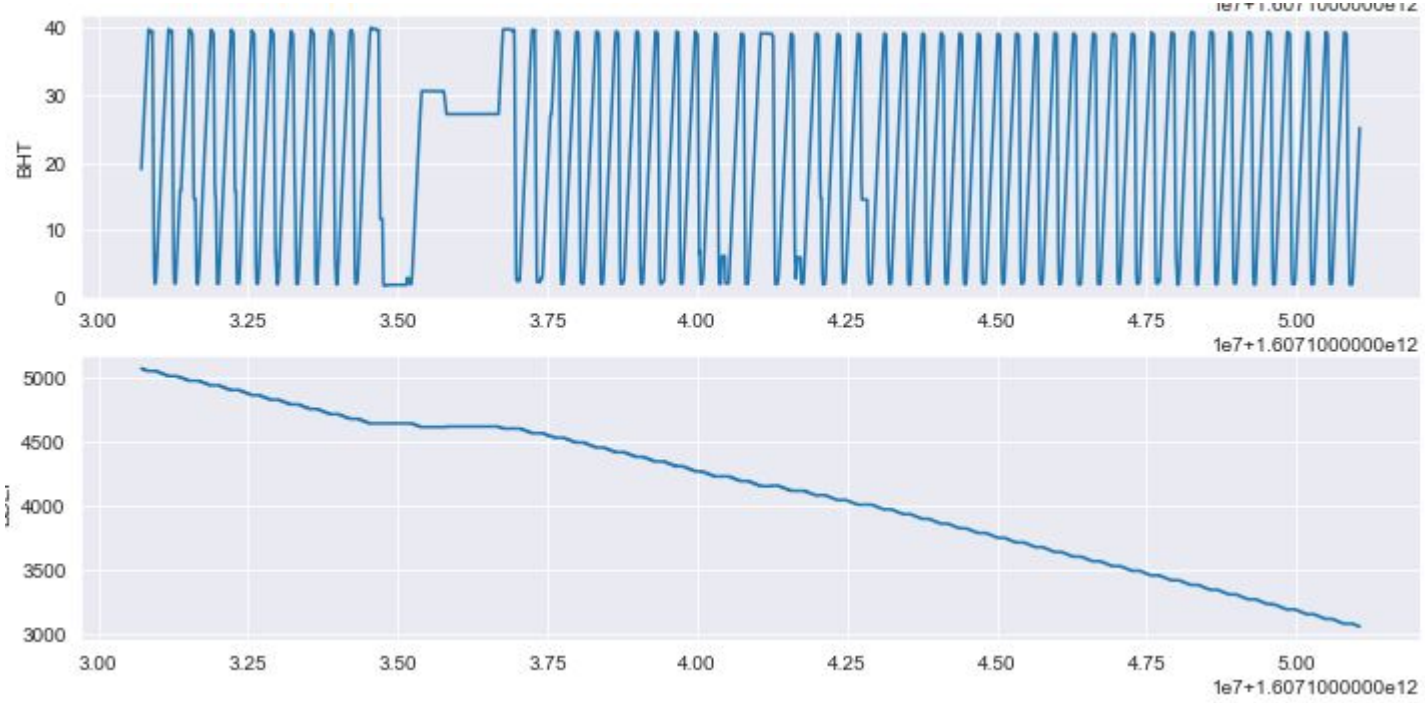
Correlation Analysis



Time-Dependent Variations



Time-Dependent Variations



3.

Modelling

Using statistics and computing to make predictions.

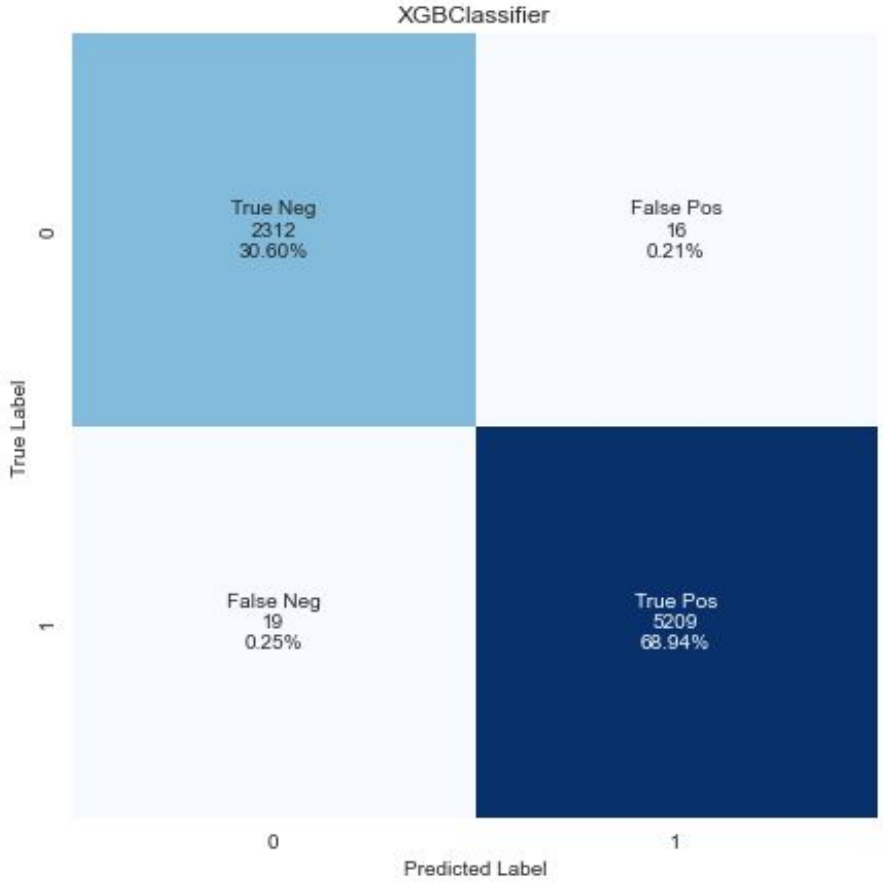
Selecting the Baseline Model

To define the baseline supervised learning model, I've evaluate the performance of the following classification algorithms:

	Model	Accuracy	Precision	Recall	F1 Score
0	RandomForestClassifier	0.991927	0.993883	0.994453	0.994168
1	LogisticRegression	0.977369	0.976806	0.990819	0.983762
2	SVC	0.976310	0.974976	0.991201	0.983022
3	GaussianNB	0.976046	0.975863	0.989862	0.982813
4	XGBClassifier	0.993912	0.995222	0.995983	0.995602
5	LGBMClassifier	0.995368	0.996938	0.996366	0.996652

Evaluating the LGBMClassifier

Let's have a look at the confusion matrix and the classification report for the selected model.



Evaluating the LGBMClassifier

Let's have a look at the confusion matrix and the classification report for the selected model.

LGBMClassifier

	precision	recall	f1-score	support
off_slips	0.99	0.99	0.99	2328
on_slips	1.00	1.00	1.00	5228
accuracy			1.00	7556
macro avg	0.99	0.99	0.99	7556
weighted avg	1.00	1.00	1.00	7556

Hyperparameter Fine Tuning

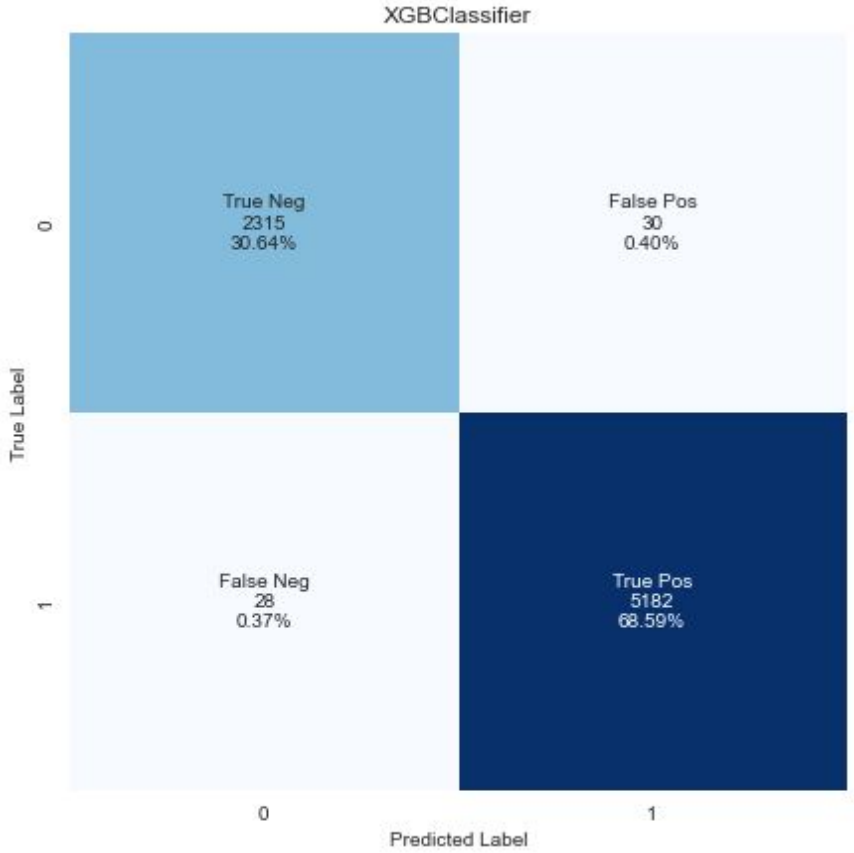
The best possible performance is the one that gives the highest f1 score

LGBMClassifier

	precision	recall	f1-score	support
off_slips	0.99	0.99	0.99	2345
on_slips	0.99	0.99	0.99	5210
accuracy			0.99	7555
macro avg	0.99	0.99	0.99	7555
weighted avg	0.99	0.99	0.99	7555

Hyperparameter Fine Tuning

The best possible performance is the one that gives the highest f1 score



4.

Using the Model

How one would use it?

Concluding Remarks

- After performing an exploratory data analysis, the statistical metrics showed us which features are the most important considering the goal of this project.
- The finds suggested that tree-based models would be a good fit for the task, which was confirmed by the classification reports.
- The supervised learning algorithm LGBMClassifier outperformed all the other tested algorithms
- The model can accurately capture the operational state of the slips with 99% precision
- The developed solution can be used for automated monitoring of the performance of drilling operations for any O&G industry with 92% of reliability on the final KPI.

Possible Next Steps

- I'd put this model in production on a cloud environment so that it can receive requests from other applications to stream predictions and deliver the wanted KPI.

- Additionally, I'd like to evaluate how this model would perform on previously unseen data.

Developing a KPI to the Slip-to-Slip Connection Time in a Drilling Operation

Pedro Grosman
grosmanph@gmail.com