

SUPERVISED LEARNING METHODS FOR DRILLING STATE MACHINE PREDICTION PROPOSAL

Prepared by Nicole McCarthy on May 29, 2025

Deere Development
Company, LLC.

Contents

Abbreviations	2
Executive Summary	3
Introduction	4
Related Work	5
Proposed Work.....	6
Evaluation	8
Discussion	8
Conclusion	9
References	10

Abbreviations

DT	Decision Tree
GPM	Gallons per Minute
JITL	Just-In-Time Learning
MFI	Mud Flow Index
MR	Motor (on a drill string)
PD	Pulser Device
RBF	Radial Basis Function (Kernel)
RF	Random Forest
ROP	Rate of Penetration
RPM	Revolutions per Minute
RT	Real Time
SVM	Support Vector Machine
WOB	Weight on Bit

Executive Summary

The client, Deere Development Company, has requested a method for predicting the drilling state machine (as defined by them) based on typical downhole drilling parameters for flow, rotation, vibration, and WOB for oil rig drilling. The drilling state machine defines five states: Pumps Off, Pumps On, Sliding Drilling, Rotating, and Rotational Drilling. By identifying the state correctly, the client can better implement other software for drilling optimization. This combination will make oil rig work more cost effective, time efficient, and safe for workers. The methods slated for development, in order of priority, are decision trees, a random forest, support vector machines, and an implementation of just-in-time learning for the SVMs.

Introduction

This project focuses on using drilling parameters of oil rig drill strings to increase the knowledge and efficiency of runs. The client, Deere Development Company, has requested a method of estimating the drilling state machine, as defined by them, based on the drilling parameters sensed by the tool in real time. The knowledge of the drilling state machine can then be utilized to turn on and off other software that optimizes drilling efficiency and safety, critically in terms of energy, cost, and time.

The drilling state machine has been defined with five modes. High and Low are relative terms to describe how each parameter changes from the last drilling state. The drilling state machine can only move consecutively with the exception of (1) providing access to (2) and (3), but (2) not providing access to (3) (see Figure 1) (Deere, 2025).

There are no existing solutions for this specific problem, but similar work has used the same drilling parameters to predict geological formations and torque on the drill string. Therefore, this project aims to expand on the methods previously used for classification in real time and shift their focus towards better understanding the drilling itself.

Drilling Parameter	Flow	Rotation	Vibration	WOB
(0) Pumps Off	Low	Low	Low	Low
(1) Pumps On	High	Low	High	Low
(2) Sliding Drilling	High	Low	High	High
(3) Rotating	High	High	High	Low
(4) Rotational Drilling	High	High	High	High

DRILLING STATE MACHINE

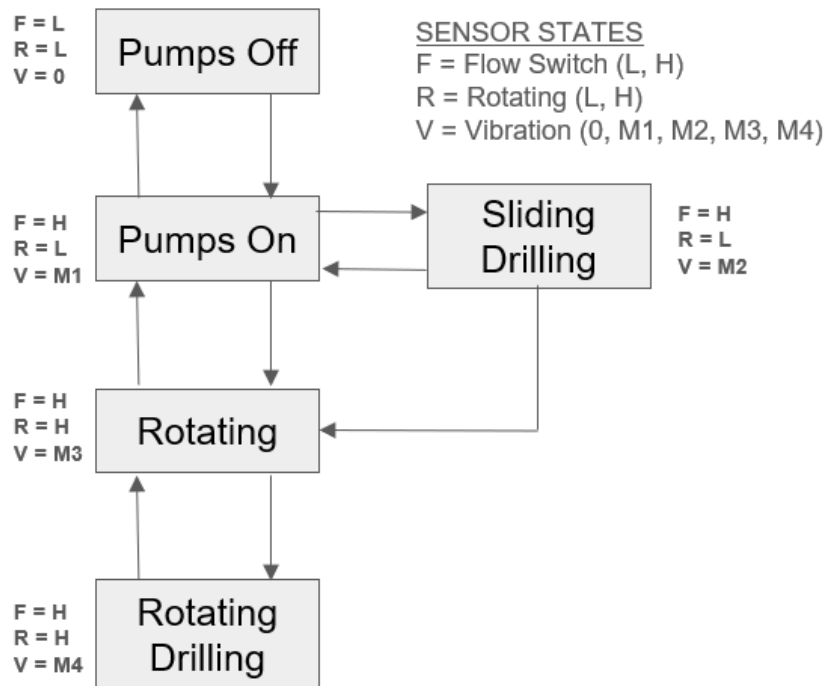


Figure 1 Deere Development Company's Drilling State Machine

Related Work

The problem of using drilling parameters to predict things about the downhole environment has become popular recently, especially for predicting geological formations. These projects have similar parameters to those proposed in this project to solve other problems. For instance, Hassan et al. (2024) tested three different models using ROP, GPM (flow), RPM, strokes per minute, torque, and WOB to predict the porosity and permeability of the geological formations being drilled. Of decision tree (DT), random forest (RF), and support vector machine (SVM) models, each had a correlation coefficient of above 0.8, but RF performed the best with 0.92. This is likely due to the robustness built into RF, since it uses the aggregate of multiple decision trees. A visualization on the prior work done on predicting geological formations from

drilling parameters can be found in Hassan et al. (2024, p. 17068). These methods range from supervised to unsupervised learning.

Similarly, Bai et al. (2025) found that a just-in-time learning (JITL) model outperformed a sliding window model in predicting the torque downhole due to its ability to draw from historical data similar to the parameter data the tool was finding in real time. The parameters used in this experiment included time, depth, WOB, hook load, RPM, MFI (flow), and surface torque. Euclidean distance, cosine similarity, and an SVM were used to identify historical data that was most similar to the real time data.

This project will build on the findings of these experiments by focusing on supervised learning techniques and applying the drilling parameters to predicting the drilling state machine rather than the surrounding environment or torque. This technology will be more applicable to a wider range of associated technologies for optimizing drilling in terms of efficiency, safety, and accuracy.

Proposed Work

The datasets for this project have been provided by the client, Deere Development Company. They represent six runs from three different oil rigs, named Flybar 1WB, Flybar 1WC, and Flybar 2WC. These datasets include time and depth as well as the following attributes and validators relevant to this project. Each item indicated below represents a column in the dataset that will be used for the project. RT refers to “real time” at the surface, PD to the pulser near the drill bit, and MR to the motor which is just above the bit.

DRILLING PARAMETER	VALIDATORS	DATA
FLOW	<ul style="list-style-type: none"> MR_Flow.M PD_Pump_Event 	<ul style="list-style-type: none"> RT_Pumps_Off, RT_Pumps_On
ROTATION	<ul style="list-style-type: none"> MR_ROTATION 	<ul style="list-style-type: none"> MR_RPM-AVG.MR RT_RPM.W
VIBRATION	<ul style="list-style-type: none"> RT_SLIDE_BIT* 	<ul style="list-style-type: none"> MR_VIBA, MR_VIBL PD_Axial Vibration, PD_Lateral Vibration
WOB	N/A	<ul style="list-style-type: none"> RT_WOB

*This validator is not for vibration specifically, but whether the tool is sliding, which would have different levels of vibrations detected.

Each of the validators have a 1/0 where the indicated parameter is turned on or off, respectively. The two data columns that represent flow (in GPM), will be aggregated together as they almost never coincide in the same row.

This project will focus first on creating ten decision trees that will be incorporated into a random forest and then on establishing a JITL model for SVMs. These three methods were chosen for their easy interpretability, efficiency, robustness, and their accuracy in previous work.

The ten DTs will each include WOB, and the flow aggregate mentioned earlier. The variation will come from a unique pairing between each dataset for rotation and vibration, as well as an average of both options. The tenth DT will include all options for rotation and vibration. Finally, the RF will aggregate the outcomes from each DT and settle on the most popular one.

The JITL for SVMs will test different kernels for flow, rotation, vibration, and WOB, including linear, polynomial, and radial basis function (RBF), to determine the best ones for each variable. If there is enough time at the end of experimentation, then an ensemble approach may be utilized with the best performing DTs (or the RF) and SVMs.

Evaluation

This project will utilize the data from Flybar 1WB and Flybar 1WC to train the three methods. These methods will then be tested on the singular, but lengthy, run conducted on Flybar 2WC. A comparison between each DT, the RF, and the JITL method will be the final focus of this work. The metrics for evaluation will include mean absolute error and accuracy (including a confusion matrix).

Discussion

The current status of the project is in the preprocessing stage. A lot of time has already been spent getting to know and understand the data. Some preprocessing has already occurred, including getting the data into individual runs and removing unnecessary information. The final portion of preprocessing left is getting the remaining data into a usable format for the methodologies proposed. A rough timeline for the rest of the project is below.

The most significant challenge for this project is the condensed timeline. The timeline for the project will be re-evaluated at the Project Checkpoint. Should the scope of the project prove to be too large for the given time, the JITL portion of the SVM method will be eliminated and the SVMs will be evaluated without the additional optimization. Even if this proves to be too much, the several DTs and RF will be enough for a comparison of the most useful datatypes for implementation. At this time, there are no other challenges of significance anticipated.

DATE	ASSIGNMENT DUE	ASSIGNMENT TASKS	RESEARCH TASKS
JUNE 2			Preprocessing
JUNE 3			Build DTs
JUNE 4			Build RF
JUNE 5			Test DTs and RF
JUNE 6			Build SVMs
JUNE 9		Current Work	
JUNE 10		Evaluation and Discussion	
JUNE 11	Project Checkpoint	Summary and Conclusion	
JUNE 12			Test Kernels
JUNE 13			Create JITL Querying
JUNE 16			Test JITL System
JUNE 17			Evaluate All Methods
JUNE 18		Update Introduction, Related Work, and Methodology	
JUNE 19		Results	
JUNE 20		Discussion and Conclusion	
JUNE 23	Final Report	Summary and Presentation Slides	
JUNE 24	Final Presentation	Record Presentation	

Conclusion

This is an ambitious project taking course over the month of June 2025. Its goal is to create and evaluate methods of predicting Deere Development Company's drilling state machine using drilling parameters for flow, rotation, vibration, and WOB. By using the most accurate and efficient methods, chosen based on previous work for similar applications, the client will be able to implement other software with higher effectiveness and significantly decrease costs, time, and safety risks for workers. The building and testing of methods will be prioritized as such:

1. 10 Decision Trees (each with slightly different parameter inputs)

2. Random Forest of previous DTs
3. Support Vector Machines for Flow, Rotation, Vibration, and WOB (combined into one drilling state)
4. Just-In-Time Learning application to SVMs (for increasing efficiency and accuracy)

The methods outlined above will be trained on drilling data from five runs on two different wells. It will then be tested on one run from a third well. All drilling data was provided by the client from real oil rig work. Once the methods that were developed (based on time availability) have been tested, they'll be evaluated for performance based on error and accuracy.

References

- Bai, K., Sheng, M., Zhang, H., Fan, H., Pan, S. (2025). "Real-Time Drilling Torque Prediction Ahead of the Bit with Just-In-Time Learning." *Petroleum Science* 22, 430-441. <https://doi.org/10.1016/j.petsci.2024.12.014>.
- Deere, P. Personal Communication. March 8, 2025.
- Hassaan, S., Mohamed, A., Ibrahim, A. F., & Elkatatny, S. (2024). "Real-Time Prediction of Petrophysical Properties Using Machine Learning Based on Drilling Parameters." *ACS Omega*, 9(15), 17066–17075. <https://doi.org/10.1021/acsomega.3c08795>.