Achieving Optimal Horizontal Drill Operations

Daniel Serna[1], James Vasquez[1,2], and Donald Markley[2]

[1] Southern Methodist University, Dallas TX 75275 USA
{dserna,vasquezj}@smu.edu
[2] Triple Crown Resources, 1722 Routh St. Suite 1750, Dallas TX 75201 USA
dmarkley@triplecrownresources.com

Abstract. In this paper we present a novel method of predicting when a slide event will occur in horizontal drilling operations. A slide event occurs when a drill stops efficiently rotating through the subsurface. Rate of penetration is significantly reduced during a slide event. We achieve this novel method by creating a predictive model based on a number of features present in time series data provided by Triple Crown Resources.

1 Introduction

Drilling a well involves many moving parts to reach total depth of the well. While drilling in a horizontal fashion, drill sensors record the rate of penetration. The higher the rate of penetration, the quicker the drill penetrates through the subsurface rock. A stable and higher rate of penetration will allow the drilling operations to reach total depth faster. Total depth is reached fastest when the drill is in a state of rotation as this provides the greatest rate of penetration. However, when the drill is in a sliding state, rate of penetration is significantly reduced.

The team has created a predictive model that minimizes the mean squared error in predicted versus actual slide events. The team's model predicts an impending slide event within a specified time frame.

During drilling operations, slide events are identified by monitoring the rate of penetration, bit inclination, and pipe consumption. When a slide event is detected, drill operators use the ROCKit system for adjustments. With this system, the drill pipe is rocked back and forth in an effort to return the drill back to a rotating state[1]. Using our predictive model, drill operators can employ the ROCKit system or other methods prior to a slide event occurring in order to prevent the drill from ever entering a sliding state.

Triple Crown Resources's budget for capital expenditures is dominated by the expenses associated with drilling and hydrocarbon production. By reducing capital spend in this area, Triple Crown Resources would be able to obtain larger investments and expand their drilling operations to surrounding areas. Furthermore, by reducing the time needed to reach total depth, Triple Crown Resources

[1] Nabors. "ROCKit System", https://www.nabors.com/software/performance-drilling-software/rockit-system [Accessed 13 July 2019]

would be signalling to equity partners that they have mastered advanced analytic techniques allowing them to perform better than their competitors.

Our paper first provides context around the problem domain in Section 2.1, specifically explaining what a slide event is. We then provide an explanation of horizontal drilling operations and a bit of history regarding earlier drilling techniques in Sections 2.2 and 2.3. Section 2.4 provides an explanation as to why there is such interest in the Permian Basin currently for horizontal drilling operations. An explanation of our data and cleansing techniques follows in Section 3.1 with an brief discussion of our exploratory data analysis in Section 3.2. Section 4 explains why we chose certain features for our model. Sections 5, 6, and 7 discuss our application of Support Vector Machine, Random Forest, and Logistic Regression models to the data. We then discuss our results in Section 8. We point out ethical considerations in Section 9 and lay out our conclusions in Section 10.

2 Drilling and Slide Events

As drilling technology advanced, the need to optimize drill operations to prevent slide events became more important. While optimal drilling parameters have always been a concern, slide events are less prevalent in vertical drilling operations. As horizontal drilling operations became more popular, the focus on preventing slide events increased.

2.1 Slide Events

A slide event occurs when the rate of penetration of the rotary drill decreases and the drill bit no longer rotates at optimal conditions. Ideally, rotation is the desired state of a drill bit. One can think of efficient drill rotation as similar to how a screw rotates through a piece of wood. When a drill is sliding, it is pushing through the rock formation instead of efficiently rotating. Slide events can occur due to the type of rock formation encountered by the drill bit, the bit condition, and the amount of formation pressure encountered.

2.2 Horizontal Drilling

A horizontal well is constructed such that it can make a turn in the subsurface usually at 75 or 85 degrees. In Fig. 1, well paths labeled "H" are considered horizontal wells whereas well paths labeled "V" are considered vertical wells. The key difference among these two wells is well paths labeled "V" never reach an inclination of 75 degrees or greater, and as such are not considered a horizontal well. Horizontal drills operate in order to maximize reservoir extraction by allowing the lateral portion of the well to stay within the producing rock formation longer.

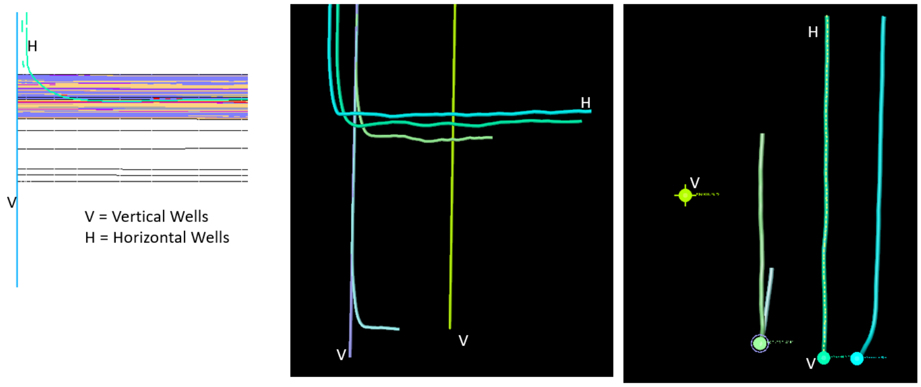

Fig. 1. Types of Well Paths: Vertical wells were the dominant type early in the industry as technology and understanding of reservoirs were not as advanced as they are today. Horizontal wells gained acceptance in the early 2000's and are now a standard type for drilling operations.

2.3 Drilling Advances

A traditional vertical well contains a bore hole that extends vertically below the derrick. In a true horizontal well, the bore hole goes from vertical to horizontal at a point in the well path (see Fig. 1). The first vertical well was drilled in 1895 using a percussion drilling method and reached a depth of sixty five feet. For comparison, modern day deep water wells have reached depths of 24,000 feet. Technology advancements led to a new type of drilling technique called rotary drilling. The basic concepts for rotary drilling brought forth in the late 1800's can still be seen in today's operations as drill pipes spin as they drill down into the subsurface. This type of drilling operation is a necessity for energy companies to reach deep target depths.

Directional drilling was initiated to fish tools lost in a bore hole [2]. Intentional directional drilling methods were first used in 1930 on the shores of Huntington Beach, California. In 1934, directional drilling was used to kill a well blowout by pumping heavy mud into a deviated borehole.

A leap in technology and computing power thrust many industries, including Energy, into Big Data Analytics. However, the velocity of the data captured proved to complicate analysis for the Energy industry [3]. Specifically, real time drilling sensors capture data every second. Many in the industry were not experienced in how to analyze and make the best use of this data. With the market drop in 2014, Energy companies began to realize this data was a valuable asset. Efforts were put forth to recruit talent that could make use of these large datasets that had been sitting idle in 3rd party vendor databases.

2.4 Permian Basin

The Permian Basin stretches from the lower Southern portion of New Mexico and extends to much of West Texas (see Fig. 2). This basin was formed during the Paleozoic era. From the geological timeline, much of the structures which ultimately formed the traps for hydrocarbon were created during the late Paleozoic Era (251 million years ago) [7].

The basin is divided into three structural development phases. This paper concentrates on the tectonic activity of the Hercynian Orogeny which occurs when the North America plate collides with the South American plate. This tectonic activity is one of the major factors that contributed to the conventional traps exploited by early gas and oil exploration. Eventually, as tectonic activity slowed and mountains eroded, sediment consisting of limestone, shale, and fine-grained sandstone layered the basin [6].

Oil was first produced from the basin in the middle of the 1920's, and major activity started during the 1950s. Much of the data from these early periods is still used today to deliver control points of the basin. These control points help geologists map different formations in the subsurface layers [4].

In modern development, companies now have higher quality data available and are able to divide previous grouped formations into sub-formations. The Wolfcamp formation is one such sub-formation that resulted from higher quality data. This formation is a target for most companies involved in the Permian Basin. Using today's mapping technology, this formation has been divided into three benches - upper, middle, and lower Wolfcamp.

Though the Permian Basin has been producing for more than five decades, new technology emerged that brought additional life to the basin in recent years. Hydraulic fracturing ("fracking") has been around since 1949, but it was not until the early 2000's that fracking was combined with horizontal drilling techniques. The ability to drill in a horizontal direction to stay within a formation gave an unprecedented way to drain reservoirs that were once thought to be on the decline for producing hydrocarbons (see Fig. 3).

Recently, it has been reported that the Permian Basin is the world's top oil producer. Saudi Aramco's Ghawar field produced 3.8 million barrels per day in 2018. The Energy Information Administration reports that the Permian Basin is now producing 4.2 million barrels per day compared to an average of 3.4 million barrels per day for 2018 [5]. Thus, oil production in the Basin is increasing and companies are investing more capital.

3 Data

The data provided to our team originates from drilling operations conducted by Nabors Industries. This data is streamed real-time to data stores and exposed for analysis through the MyWells portal. The data provided covers twenty one wells and is quite large. Our methods of cleansing and sampling are explained in detail in Section 3.1. Our explortory data analysis and resampling techniques

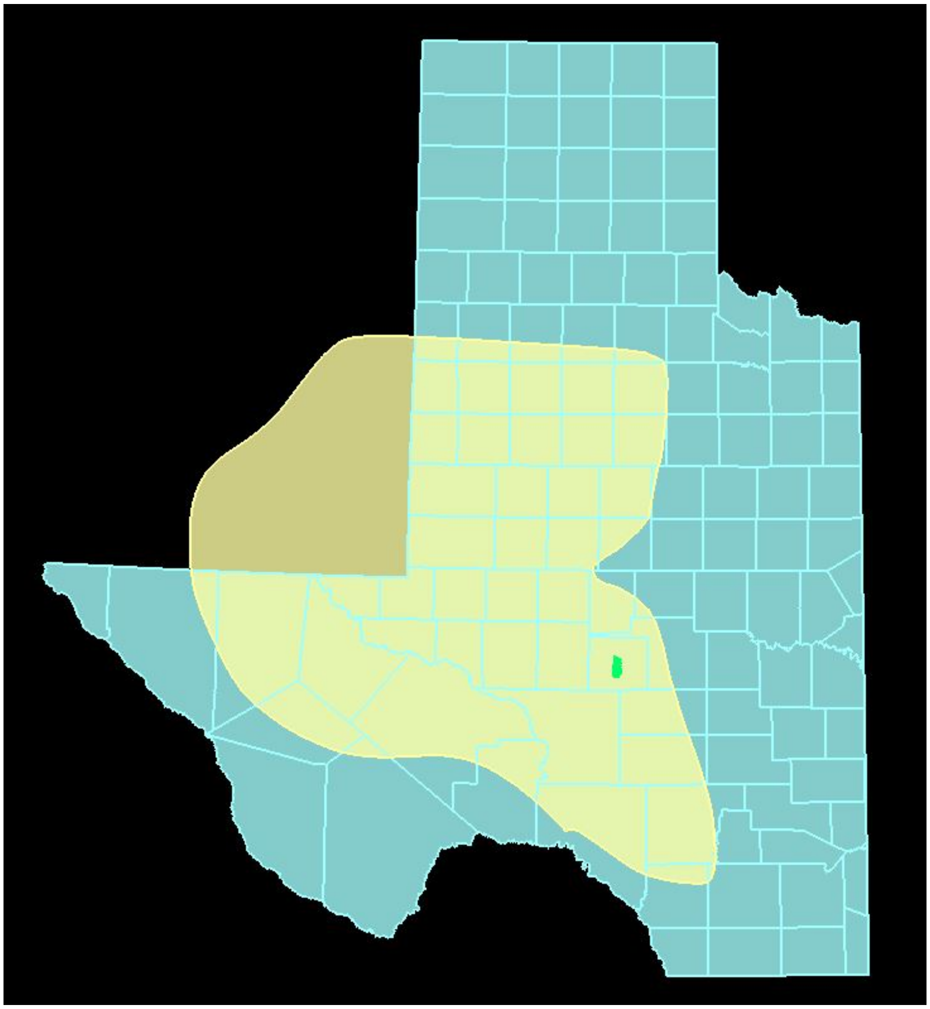

Fig. 2. Permian Basin: The Permian Basin is roughly 86,000 square miles and split into multiple sub basins. Midland, Delware, and Val Verde make up the larger Permian Basin. Each basin comes with its own unique challenges such as shale thickness, mineralogy, pressure, and access to drill sites. Formation water is an operational challenge that affects all companies in the basin. Water that is produced from a well cannot be pumped back down and requires transportation and treatment. Up to 15,000 barrels of formation water can be produced daily from a producing well.

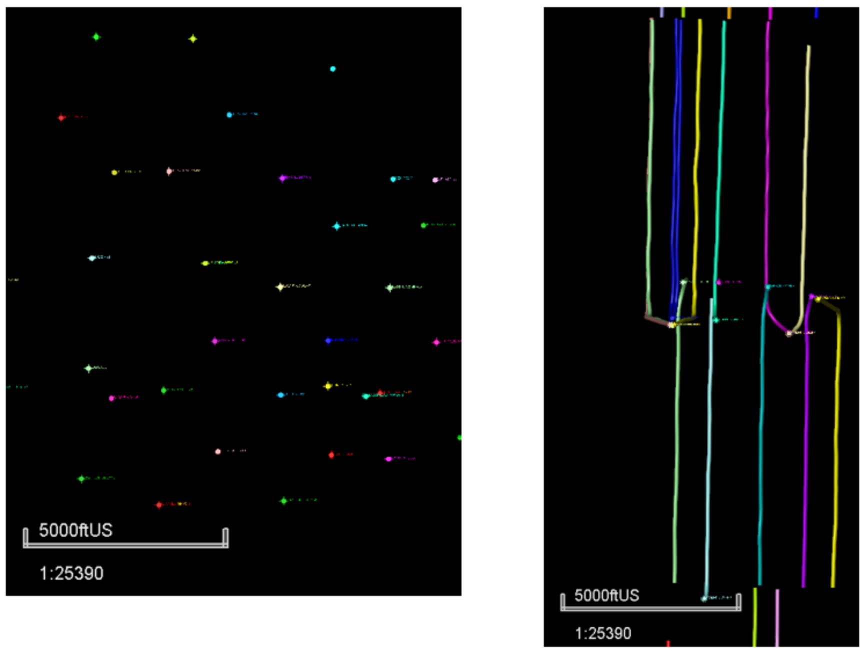

Fig. 3. Vertical versus Horizontal Drilling: Horizontal drilling programs bring a lesser environmental impact than traditional vertical wells. Draining a reservoir with vertical wells requires a much bigger environmental footprint, requires more drills, and costs more. A single pad can hold as many as eight horizontal wells.

are explained in Section 3.2. Most importantly, the data set contains a feature "Rockit-OnOff" which indicates if a drill is in a sliding or rotating state. This feature is used as our prediction variable. The Nabors system allows for monitoring of drill operations which can be used to validate the success of our model (see Fig. 4).

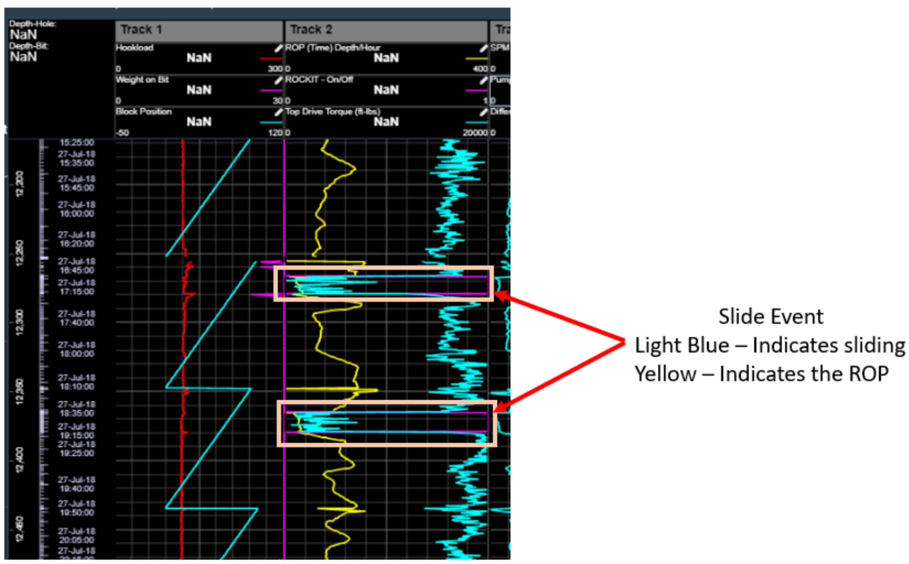

Fig. 4. MyWells Interface: A real time operational dashboard used while drilling. Drill operators use this information to control drilling operations and maintain optimal drilling parameters. The highlighted areas above shows how the rate of penetration drops significantly during a slide event.

3.1 Cleansing and Preparation

Triple Crown Resources has access to proprietary sensor data of 21 wells currently and the well count is increasing. Data points are captured every second while drilling which results in an extremely large dataset. Due to the size of the dataset, it is the team's decision to take samples of the data every 10 seconds. This decision reduced the amount of data to 2.98 million records initially which is more feasible to consume.

The dataset initially contained 506 features. The team decided to reduce the number of features to further reduce the size of the dataset. The team removed columns only containing NULL values or single values. In addition, the team removed columns that contained less than 90% of filled values. These decisions reduced the feature count for initial analysis down to 122.

In addition, by utilizing industry specific business insight, our team was able to further reduce the size of our dataset. Triple Crown Resources indicates records with an inclination value greater than or equal to 85 degrees is relevant to our problem domain. This inclination value indicates the record is related to the horizontal section of the well which is the focus of our problem statement. Removing records not meeting this criteria reduced our initial dataset to 725,000 records.

3.2 Exploratory Data Analysis

The team performed exploratory data analysis on the original dataset as well as the reduced feature dataset. Fortunately, the data is populated well. Only two features have notable NA percentages. MWDContinuousInclination has an NA percentage of 3.25% and MWDGammaAPI has an NA percentage of 1.66%. These features turn out not to be a concern as they were excluded from our reduced feature dataset used for analysis.

With our reduced feature dataset, none of our features have NA values. However, we do have some issues with the feature distribution. ROPMinuteDepth, Depth_Bit, ADROPSetpointValue, and DepthHoleTVD are all right-skewed. Also, TopDriveTorqueCommand shows significant deviation from normality. Since we have a large sample size, the central limit theorem should come into play and the deviations from normality are not a concern.

The team additionally chose to experiment with data re-sampling to decrease the time required to run analysis. The team attempted re-sampling windows of one minute, five minutes, and ten minutes. The results of this analysis are seen in Fig. 5. The team experienced only minor drops in model performance at the one minute re-sampling rate. Re-sampling rates of five minutes and ten minutes resulted in significant decrease of model performance. As such, the team decided to proceed with re-sampling of the data using a one minute re-sampling rate. This struck a good balance between speed and accuracy.

4 Feature Selection

After this data cleansing was performed, the team applied a Decision Tree feature selection process to select the most relevant features. Ultimately our chosen features explained 97% of the variance. Table 1 describes features commonly used for analysis, however, we did not use all common features in the model.

Table 1: Commonly Used Features

Variable	**Calculation/Description**
ADDifferentialPressureActualValue	Actual Value of the Differential Pressure based on PLC calculation

ADDifferentialPressureSetpointValue	The setpoint value entered on the HMI for Differental Pressure
ADROPActualValue	Actual Value of the ROP based on PLC calculation
ADROPSetpointValue	The setpoint value entered on the HMI for ROP
ADWOBActualValue	Actual Value of the WOB based on PLC calculation
ADWOBSetpointValue	The setpoint value entered on the HMI for WOB
BlockPosition	The position of the traveling blocks
CasingPressure	A term used in well-control operations, typically during the drilling or workover phases of a well, to describe the pressure in the drillpipe or tubing annulus.
Depth_Bit	The depth where the bit is actually sitting at.
DifferentialPressure	In general, a measurement of fluid force per unit area (measured in units such as pounds per square in.) subtracted from a higher measurement of fluid force per unit area.
FlowIn	Rate of flow being pushed downhole
GasTotal_units	Gathering of qualitative and semi-quantitative data from hydrocarbon gas detectors that record the level of natural gas brought up in the mud.
Hookload	The total force pulling down on the hook. This total force includes the weight of the drillstring in air, the drill collars and any ancillary equipment, reduced by any force that tends to reduce that weight.
Pit01Volume	Total Volume in Pit01
Pit02Volume	Total Volume in Pit02
Pit03Volume	Total Volume in Pit03
Pit04Volume	Total Volume in Pit04
Pit05Volume	Total Volume in Pit 05
Pit06Volume	Total Volume in Pit06
Pit07Volume	Total Volume in Pit07
PumpPressure	The pressure sensor is used to monitor Mud Pump Pressure.
ReturnFlow	Rate of flow on the flowline

ROPDepthHour	ROP is calculated based on a combination of depth step and time. Based on the drilling speed, the time taken for a new depth step may vary. The ROP may be calculated at different time intervals. In order to avoid this, if a new depth step is not reached within the time taken for the previous ROP calculation, a new ROP will be calculated immediately. In this case the ROP won't be calculated at every depth step, but only based on previous ROP calculation time. ROP is calculated based on the change in total depth during that time interval.
SPM1	The number of strokes the polished rod completes in one minute. This determines the rate at which liquid is pumped. If the number of strokes per minute is increased, the pump rate is also increased.
SPM2	The number of strokes the polished rod completes in one minute. This determines the rate at which liquid is pumped. If the number of strokes per minute is increased, the pump rate is also increased.
SPMTotal	Combined SPM for all pumps
Strokes1	A count of stokes on pump 1
Strokes2	A count of stokes on pump 2
StrokesTotal	Count of strokes for all pumps
TOP DRIVERPM	Amount of RPMs generated from the Top Drive
TOP DRIVETorqueft_lbs	Amount of Torque generated from the Top Drive
WeightonBit	Amount of weight on the bit

5 Support Vector Machine

Background of SVM Model....placeholder text.

The team applied a Support Vector Machine model against the data using a Radial Basis Function kernel. With the Support Vector Machine model, the team achieved an accuracy of 0.98966, precision of 0.97791, recall of 0.98811, and an F1 score of 0.98298.

6 Random Forest

Background of Random Forest Model...placeholder text.

The team applied Random Forest against the data using 10 estimators. With the Random Forest model, the team achieved an accuracy of 0.97823, precision of 0.97612, recall of 0.95127, and an F1 score of 0.96353.

	Raw			
Model Type	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	0.98639	0.97685	0.97816	0.97751
Random Forest	0.97823	0.97612	0.95127	0.96353
Support Vector Machine	0.98966	0.97791	0.98811	0.98298
	1 Minute			
Logistic Regression	0.97578	0.95408	0.95776	0.95591
Random Forest	0.95285	0.91211	0.91675	0.91443
Support Vector Machine	0.98004	0.95853	0.96932	0.96389
	5 min			
Logistic Regression	0.92081	0.86647	0.84346	0.85481
Random Forest	0.89532	0.86584	0.73515	0.79516
Support Vector Machine	0.9241	0.86704	0.85674	0.86186
	10 min			
Logistic Regression	0.8622	0.7765	0.71409	0.74399
Random Forest	0.82009	0.75185	0.53491	0.62509
Support Vector Machine	0.86183	0.78947	0.69169	0.73735

Fig. 5. Re-sampling: Re-sampling rates of one minute, five minutes, and ten minutes were attempted. Notable decreases in model performance were experienced with five minute and ten minute re-sampling rates. However, a one minute re-sampling rate resulted in only minor decreases in model performance.

7 Logistic Regression

Background of Logistic Regression Model...placeholder text.

The team applied Logistic Regression against the data using a limited memory Broyden-Fletcher-Goldfarb-Shanno solver. With the Logistic Regression model, the team achieved an accuracy of 0.98639, precision of 0.97685, recall of 0.97816, and an F1 score of 0.97751.

8 Results and Analysis

The team achieved the best model performance (as determined by F1 score) with the Support Vector Machine model (see Fig. 6).

Model Type	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.98639	0.97685	0.97816	0.97751
Random Forest	0.97823	0.97612	0.95127	0.96353
Support Vector Machine	**0.98966**	**0.97791**	**0.98811**	**0.98298**

Fig. 6. Model Results: The Support Vector Machine model performed the best as determined by F1 score. Logistic Regression was a close second while Random Forest performed the worst.

Using the Support Vector Machine model, the team's model only misclassified five events (see Fig. 7). It appears the model is misclassifying the beginning and end of a slide event.

Our team is not the first to be interested in optimizing drill operations. Understandably, there is prior work related to our problem domain. US Patent 6,152,246 describes a database application designed to measure drilling parameters, apply operating limits, and alert operators when specified events occur [1]. While this work is intriguing, this invention is geared towards real time event monitoring and alerts. Our slide event prediction model achieves additional benefit by being able to alert operators of an upcoming slide event. This allows operators to take preventative measures before the event ever occurs.

9 Ethics

Creating a more efficient drilling program does allow for monetary gains for the industry, however the environmental impact cannot be ignored. Byproducts of drilling include large amounts of formation water that require treatment and cannot be used for irrigation due to chemical contamination. In addition, earthquakes have been on the increase in areas where fracking operations are

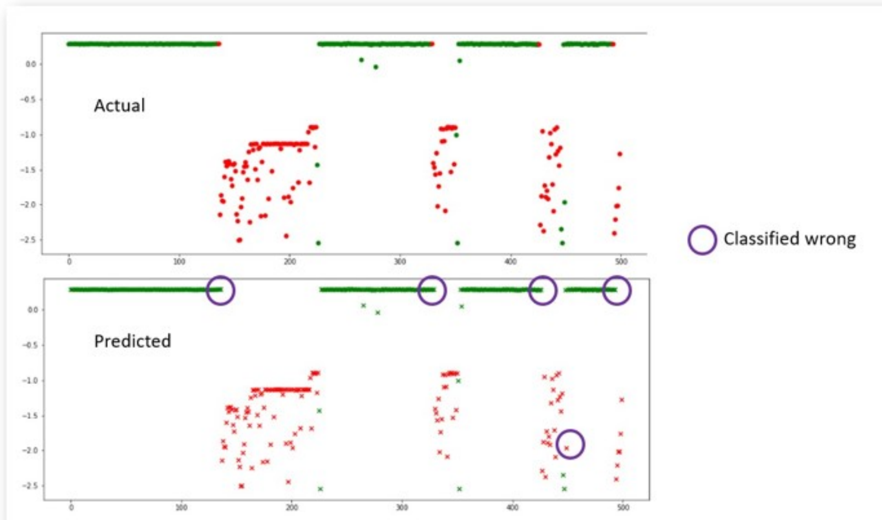

Fig. 7. Classification Results: Five events were misclassified using the Support Vector Machine model. It appears the model is misclassifying the beginning and end of a slide event.

conducted. Weighing the environmental costs against shareholder return is an ethical issue that must be discussed. The improved drilling efficiency our model brings could result in expanded drilling operations, and thus more environmental concerns.

10 Conclusions

placeholder text.

References

1. Charles H. King, Mitchell D. Pinckard, D.P.S.A.O.D.W.: Method of and system for monitoring drilling parameters. (11 2000), http://patft.uspto.gov, patent Number - 6,152,246
2. Ma, T., Chen, P., Zhao, J.: Overview on vertical and directional drilling technologies for the exploration and exploitation of deep petroleum resources. Geomechanics and Geophysics for Geo-Energy and Geo-Resources **2**(4), 365–395 (Dec 2016). https://doi.org/10.1007/s40948-016-0038-y, https://doi.org/10.1007/s40948-016-0038-y
3. Mohammadpoor, M., Torabi, F.: Big data analytics in oil and gas industry: An emerging trend. Petroleum (2018). https://doi.org/https://doi.org/10.1016/j.petlm.2018.11.001, http://www.sciencedirect.com/science/article/pii/S2405656118301421

4. Rapier, R.: Fracking has been around since 1949, why the recent controversy? Global Energy Initiative (12 2014), globalenergyinitiative.org/insights/58-fracking-has-been-around-since-1949-why, accessed 3 June 2019
5. Rapier, R.: The permian basin is now the world's top oil producer. Forbes (4 2019), https://www.forbes.com/sites/rrapier/2019/04/05/the-permian-basin-is-now-the-worlds-top-oil-producer/#6534487d3eff, accessed 8 July 2019
6. Sepmstrata.org: Permian basin tectonics - sepm strata. SEPM Strata (2019), http://www.sepmstrata.org/page.aspx?pageid=137, accessed 8 July 2019
7. Tang, C.M.: Permian basin. Encyclopedia Britannica (5 2015), www.britannica.com/place/Permian-Basin, accessed 3 June 2019

Appendix