**Problem Statement:**

Determine the probability a slide event will occur within a certain time frame.

**Background (do not use this title name) (History / Framework / Environment)**

The Permian Basin stretches from lower the Southern portion of New Mexico and extends to much of West Texas. In geological terms the basin was formed during the Paleozoic Era which during this time range from shallow seas to vibrant oceans. From the geological timeline much of the structures which ultimately formed the traps for hydrocarbon were created during the late Paleozoic Era [Brit Refer] (251 million years ago).

First oil produced from the basin started in the middle of 1920’s, the first major activity during the basins life was during the 1950s [Global Refer]. Much of this data is still used today to deliver control points of the basin. These control points help geologist map different formations in the subsurface

Though the basin has been producing for more than five decades a technology that has evolved brought additional life into the basin. Hydraulic fracturing has been around since 1949 but it wasn’t till the early 2000’s that with a combination of horizontal drilling and fracking that an old technology brought a basin back to life. 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 a decline for producing hydrocarbons.

A leap in technology and computer power leaped many industries including Energy into Big Data Analytics. However, it was the velocity of data being captured that complicated items [SD refer]. Specifically, when monitoring real time drilling sensors are capturing data every second. Many in the industry were not experience in how to analyze and make the best use of this data. With the market drop in 2014 it was first seen that Energy companies starting to treat data as an asset and put effort and recruit talent that could make the most out of data that has been sitting in 3rd party vendor databases.

**Introduction (Preface / Lead In)**

Drilling a well has many moving parts in order to reach total depth (TD) of the well. Triple Crown Resources (TCR) has requested to gain insight in an issue that has the potential to reduce the time to reach TD. While drilling in a horizontal fashion the sensors on the drill record the rate of penetration (ROP), the higher the ROP the faster it drill through the subsurface rock. A stable and higher ROP will allow the drilling operations reach TD faster. The current issue is that the drill reduces ROP or slides instead of rotating through the subsurface rock. The sliding causes a dramatic decrease in ROP, which ultimately takes longer to reach TD when dealing with slide events while drilling.

**Data**

TCR has access to all sensor data of 21 wells and growing. The data is captured at ever second while drilling, due to this the amount of data is very massive. It was the team decision to take samples of the data every 10 seconds, which reduced the amount of rows to 2.98 million rows of data. In addition to this the data contained 506 features.

Data reduction was the first step into limiting the data, using SpotFire a visualization tool the team removed columns that had NULL values, columns with only 1 number in it, and took a cut off of 90% data filled in. These items reduced the features to 122 columns for initial analysis. Working with TCR it was indicated that rows of data could be limited to rows with an inclination 85 degrees or higher. This inclination indicates that this is the horizontal portion of the well and not the vertical section. This business knowledge reduced the initial data set to 725K rows of data.

References

<https://www.britannica.com/place/Permian-Basin>

<http://globalenergyinitiative.org/insights/58-fracking-has-been-around-since-1949-why->

<https://www.sciencedirect.com/science/article/pii/S2405656118301421>