

Data Analytics and Machine Learning in the Casing and Tubular Running Services Industry

Marcus Savini and Brennan Domec, Ph.D., P.E., Frank's International

Copyright 2019, AADE

This paper was prepared for presentation at the 2019 AADE National Technical Conference and Exhibition held at the Hilton Denver City Center, Denver, Colorado, April 9-10, 2019. This conference is sponsored by the American Association of Drilling Engineers. The information presented in this paper does not reflect any position, claim or endorsement made or implied by the American Association of Drilling Engineers, their officers or members. Questions concerning the content of this paper should be directed to the individual(s) listed as author(s) of this work.

Abstract

As the Oil and Gas industry moves towards digitization and data-driven operations, users are faced with new challenges—mainly, effectively presenting and acting upon the wealth of data generated. During most discussions surrounding this topic one will almost certainly encounter the terms “Data Analytics” or “Machine Learning.” But what are these technologies, and how might one leverage their power to promote well integrity, reliability, and efficiency? This paper will address these and other related questions. Going beyond simple definitions, this paper will cite real-world Oil and Gas examples of data analytics and harnessing large sets of data to implement and deploy machine learning algorithms aimed at improving operational efficiency, connection integrity, and well integrity. This paper will also explore lessons learned, best practices, techniques, and deployment strategies for continuous improvement.

Introduction

In recent years the world has seen a massive increase in the collection of data. Everything we do in our daily lives generates data that are captured and stored. At our current pace, we are generating more than 2.5 quintillion bytes of data per day¹. The rapid increases in data collection have created opportunities and growth within the technology industry that are here to stay. Businesses are adjusting to these changes and have begun utilizing the data to make better data-driven business decisions.

Oil and Gas companies are no exception to this increase in data collection and have been generating massive amounts of data for decades. These data come in multiple forms: downhole sensor data, personnel health and safety information, video streams, inspection reports, drilling data, etc. Oil and Gas service companies have also been collecting data on services rendered. As the industry evolves, there has been a shift from merely collecting data to effectively collecting and utilizing the data for business gain.

Data Analytics and Machine Learning

Data Analytics, in its simplest definition, refers to any method or technique applied to collected data with the goal of discovering useful information. But, collecting the data is only the beginning of the process. Additional steps may be required to make the data useful. This includes data cleansing, data

transformation, and data visualization, to name a few. In the statistical domain, additional steps may also include exploratory and confirmatory data analysis, as well as descriptive statistics. All of these steps share commonality in that they seek to make the data more useful in the quest to learn more about what has been collected. It is not uncommon to use multiple methods to gain the full potential from the data.

Machine learning (ML) is a subset of artificial intelligence (AI) that can learn from and make predictions on data. Machine learning algorithms typically require a large set of data that can be used to teach or validate an algorithm. The method of teaching can be generally divided into two main groups: supervised learning and unsupervised learning. Supervised learning requires a known output for each unique input and learns to predict the output through iterative training. Unsupervised learning requires only input data and is used to find commonalities in the data through grouping of common features. This is particularly useful when output data are not known and finds application to problems such as anomaly detection. There are dozens of machine learning algorithms in use today, including Neural Networks, Decision Trees, Support Vector Machines. Which algorithm is best suited to solving a given problem requires good understanding of the problem at hand. It is not uncommon for multiple algorithms or even custom algorithms to be required.

Casing and Tubular Running Services

Casing and tubular running services in the Oil and Gas industry consists of the handling and installation (or removal) of tubular strings. These strings are used at various stages of the well construction process and perform specific functions. One example is a completion string, which is installed into a previously drilled wellbore to complete the well and produce hydrocarbons. The tubulars are connected end-to-end, either directly or via a coupling. As the string is deployed downhole, additional tubulars are added to the string until the pre-defined length or depth is reached. Casing and tubular strings are a vital part of the well construction process, and the processes and parameters of joining the casing or tubulars together is critical to the integrity of the well.

The process of connecting tubulars together is referred to as “making-up” the tubulars. This typically requires a top rotating device and a bottom gripping device to rotate the top tubular

relative to the bottom until the appropriate make-up criteria are achieved (*e.g.*, torque, turns, delta turns, or delta torque). During make-up, data are acquired that represent the turns of the connection and the corresponding torque value. Upon completion, these data are displayed to the operator, generally as a plot of torque versus turns (*Figure 1*). He or she then analyzes the graph and determines final disposition of the connection; *i.e.*, accept the connection or reject it. If the connection is rejected, the operator must determine the cause of the rejection and prescribe a remedy to correct it. Existing systems rely heavily upon operator training to properly make-up connections and to analyze the data for any anomalies that may have occurred in the process. As a result, the accuracy and reliability of the final connection analysis is somewhat subjective.

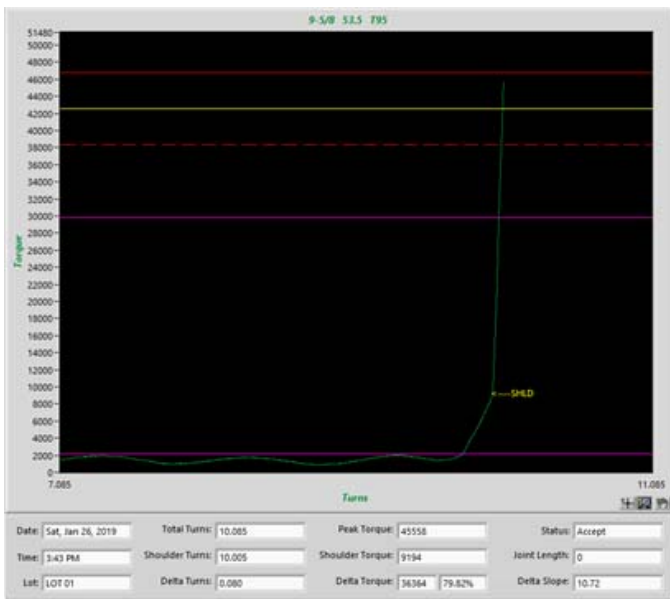


Figure 1 – Torque vs Turns make-up graph

Data Analytics for Casing and Tubular Running

The use-cases for data analytics in the casing and tubular running industry continue to grow with new demand to improve efficiency and reliability. In a recent example, data analytics processes were applied to tubular connection make-up data. The goal was to extract useful information from the dataset that could lead to improvements in efficiency and reliability as well as identification of trends and patterns not previously detected. To achieve these goals, the data analysis project was divided into five stages²:

- Define Objectives – Define clear and concise objectives for the project. This can typically come in the form of a list of questions from which conclusions can be drawn by exploring, analyzing, and visualizing the data.
- Understand the Data – Understand the data available

including the volume, variety, velocity, and veracity.

- Data Preparation – Data cleansing and transformation techniques applied to facilitate the discovery of useful information in the data.
- Data Analytics and Visualization – Visualize and explore the data for trends and patterns to reach conclusions.
- Deployment – Apply insights to operational and business processes.

Define Objectives

The start of a data analytics project should always begin with a clear understanding of what can be learned from the data, and the identification of key components that can be explored. As it relates to the tubular running services (TRS) industry, one can explore and draw many conclusions from data collected on a typical job. For example, Health, Safety, and Environment (HSE) data can be used to identify safety concerns whilst job debriefing data can be used to find trends and patterns leading to operational challenges.

After defining the areas of focus, a simple list of questions should be created that clearly define what we intend to learn from the data. In the current example, a few of the key questions to be answered included:

- Which equipment combination provides the optimal torque control?
- What effect external factors have on torque control?
- What is the most run connection?
- What are the installation efficiency differences between deepwater, shelf, and land work?
- How does running speed vary by connection type?

Understand the Data

Once the objectives are clearly defined and understood, the next step is to find all relevant data sources and identify any missing data sources needed to complete the project. Understanding the data means to know when, where, and how the data are stored, structured, and retrieved. In addition to understanding the characteristics of the data this also involves gaining a clear understanding of how the data are captured and what that means to the process.

In the connection make-up process, the data typically contain information on the job, connection, and equipment in addition to make-up graph data. These data are captured throughout the job and are stored in a master database. Depending upon the extent of operations, this database may contain many terabytes of information spanning many years and many thousands of jobs across the globe. Whilst the data within the master database is of significant value alone, opportunities to cross-reference this information with other databases may lead to even further insight. Such databases may be internal to a given company or available in the public domain; *e.g.*, historical weather data.

Beyond understanding what the data can tell us, it is important to understand what they cannot. This stage should

also investigate weaknesses in the data and where data quality and integrity issues may arise. As is often the case, the data captured may not be in an easy to search format and may take multiple cleansing and transformation operations to convert the data into a useable format.

Data Preparation

Data preparation is a critical stage in the process to improve data quality. Within the data preparation stage is data cleansing. Cleansing is an iterative process that transforms the data, as needed, to a useable format. It is necessary to gain full value of the data available and maximize the overall benefit.

In the current example, data preparation included transformation of text into enumerated values, elimination of corrupt data, and generation of retrievable summary tables. A few of the key parameters selected and normalized to yield the best results included:

- Pipe Size
- Pipe Connection
- Pipe Weight
- Thread Compound
- Location
- Equipment IDs

Throughout the cleansing process, subject matter experts (SMEs) were consulted to ensure that the data transformations did not negatively affect the data integrity.

Data Analytics and Visualization

After the data have been adequately prepared, they may now be analyzed with the goal of answering the previously-defined questions and hypotheses. Whilst analysis may include any of a variety of computational methods, data visualizations should not be underestimated in the benefit that they provide in identification of trends and patterns. There are many visualization packages available on the market today that will readily integrate with standard database architectures, expanding their use into areas not previously considered. Of course, not all data can be readily visualized in a meaningful manner requiring these instances to be addressed on a case-by-case basis or evaluated solely using computation methods.

The prepared data were analyzed using stochastic techniques and visualized where appropriate in the current example. This allowed for insight to be gained into operational performance and trends during the connection make-up process. One example is the statistical analysis of the delta turns parameter per connection (**Figure 2**). Such information provides a comparison of field performance against the performance under ideal laboratory conditions published by the connection manufacturer. This information may also be utilized to create custom rules or feed machine learning algorithms to alert operators to anomalies.

PipeConnection	PipeSize	Count of JointID	Average of DeltaTurns	Min of DeltaTurns	Max of DeltaTurns
XTREME III	22.000	68	0.026	0.000	0.048
	10.750	28	0.023	0.013	0.033
	11.750	236	0.039	0.023	0.067
	2.375	20	0.086	0.000	0.073
	4.000	6	0.071	0.062	0.078
	6.000	92	0.066	0.047	0.088
	8.625	136	0.053	0.004	0.128
	9.875	800	0.033	0.000	0.084
Total		1511	0.038	0.000	0.128

Figure 2 – Delta Turns Data Analytics Visualization

In yet another example, the differences in run-in-hole (RIH) efficiency between geographic locations were explored (**Figure 3**). This facilitates the evaluation of such factors as training and personnel competency, operating procedures, and equipment maintenance upon overall operational efficiency. When combined with additional sources such as historical weather data, insight may also be gained into the effects of external factors upon global operations.

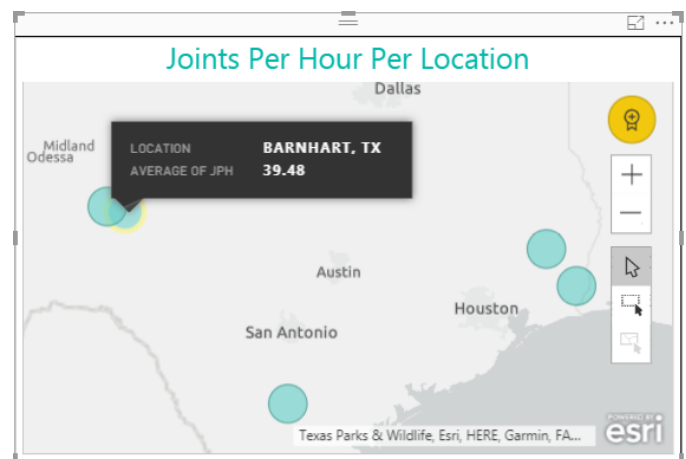


Figure 3 – RIH Efficiency Data Analytics Visualization

Deployment

Once all questions have been answered and conclusions have been drawn, the final step is to deploy the results of the analysis. Deployment may take many forms including formal reporting, creation of a dashboard, publication of a querying application, or any combination of these. The key to deployment is to deliver the information in a readily digestible and informative manner. Which form this takes will vary depending upon the problem at hand. For example, for the connection make-up scenario, a user-friendly dashboard with graphical user interface was created and deployed. This dashboard allows the user to query the data for desired relationships and drill-down in granularity as required. The dashboard also employed machine learning algorithms to provide a predictive analytics feature. This allows for new relationships to be estimated based upon the wealth of existing data within the database.

It should be understood that the preceding are just a few examples of what can be learned from a data analytics project.

In reality, the insights gained are often considered to be confidential in nature, and as such, are not published in the public domain. Generally speaking, the insights gained from the current example have allowed for improvements in design optimization, equipment selection optimization, and even forensic analysis, just to name a few. Regardless, proper deployment and targeted implementation of useable data can make the difference between a successful and an unsuccessful analytics project.

Machine Learning in Casing and Tubular Running

Completion of the data analytics project described in the preceding section had the additional benefit of answering yet another hypothesis: whether or not the data can be used to automatically evaluate and assess connection integrity. The answer is a resounding yes. Through appropriate machine learning techniques, these data were utilized to disposition, diagnose, and prescribe correction to tubular connection make-ups with higher accuracy and reliability than human operators. This section will outline the general steps followed to achieve this result.

Interest in machine learning has grown substantially in recent years and has been proven to be an effective solution for many processes within the Oil and Gas industry. In the present example, machine learning was applied to the automated evaluation of connection integrity for casing and tubulars. This application of machine learning technology followed stages similar to those discussed previously for the data analytics portion of the project. They are summarized as follows^{2,3}:

- **Define Objectives** – Define clear and concise objectives for the project. Machine learning objectives can often be defined in terms of acceptable accuracies. An ideal threshold should be set in this stage to be able to verify success of the project and to avoid unneeded iterations in the pursuit of unrealistic accuracies.
- **Understand the Data** – Understand the data available. Select parameters that may be important to solving the problem. These are called features in machine learning.
- **Data Preparation** – Data cleansing and transformation techniques applied to maximize the benefit of the data. In many cases, transformation of the data is necessary to feed the machine learning algorithm, or algorithms, the values expected. Data partitioning is also important to ensure that the data are balanced and are not prone to overtraining for a given condition. Lastly, data validation is required to ensure that the data are in fact accurate.
- **Build and Train Machine Learning Models** – Choose from multiple machine learning algorithms, or develop custom methods, into which data are fed, then choose those that yield the best results towards the stated objectives. Iteration and optimization will likely be required. In some instances, multiple algorithms may

work together to overcome limitations of each individual algorithm.

- **Deployment** – Deploy the machine learning system in a usable format. It is also necessary to plan for the future by ensuring that processes and procedures are in place for continuous improvement.

Define Objectives

A clear and concise understanding of the problem at hand is required at the beginning of a typical machine learning project. Is this a problem of regression, classification, anomaly detection, or some combination of these? What is considered a viable solution? How is success measured? What data are required? The answers to these questions directly impact success and should be well considered prior to moving forward.

For the current example, the machine learning engine is expected to correctly disposition the connection as acceptable or rejectable, diagnose the cause if rejectable, and prescribe a correction remedy. Such a problem is one of classification, and potentially anomaly detection depending upon execution. Decisions regarding accuracy thresholds were made in such a manner as to bias the machine learning system towards the conservative end of the spectrum while still balancing overall performance. This avoids incorrectly accepting connections that should be rejected; an event that could have serious implications upon well integrity.

Understand the Data

In the second step, key questions surrounding the data must be answered. These include:

- What variables or features are important to solving the problem?
- How are the data structured? How should they be structured for training? What types of transformations are required?
- How are the data stored? What retrieval methods are available?

At this stage, the selection of key parameters may be entirely based upon theoretical aspects of the process. These will be evaluated for validity and applicability later in the project. As it pertains to tubular connection make-up, the following features were initially identified as relevant:

- Graph datapoints (turns, torque, RPM, time)
- Connection make-up parameters; *e.g.*, minimum / optimum / maximum torque, connection type
- Points of interest; *e.g.*, shoulder point

Many machine learning algorithms expect values in a known or structured format and work best if inputs and outputs are scaled to range between 0 and 1. Plans for data transformation and cleansing should take into account the optimal structure and data type for the chosen algorithms to avoid iterating on a sub-optimal configuration.

Lastly, considerations should be given to easy and quick data retrieval during training and validation stages. Computational inefficiencies in this respect may induce deployment delays and are often preventable.

Data Preparation

Data preparation is a critical step for any machine learning project. During this step, data are cleansed, transformed, and validated for training. Classification algorithms often require a subset of the data that are balanced and have an equal number of potential classifications present in the sample. This avoids overtraining of the model should it be trained with data biased heavily towards one classification; *e.g.*, acceptance of the connection. Throughout the data preparation stage, SMEs play an important role to ensure the data integrity is not compromised by these transformation and sampling methods.

Build and Train Machine Learning Models

At this point the objectives are defined, and the data have been selected and prepared for use. The next step is to begin building and training a machine learning model and evaluate the results. As previously discussed, there are dozens of algorithms and techniques in common use today. Selection of the optimal technique that best solves the problem takes experience, knowledge, and/or a lot of time for trial and error.

In the current example, multiple classification and anomaly detection algorithms were selected for evaluation. The results of this evaluation are best summarized in a confusion matrix (*Figure 4*). A confusion matrix identifies where the model correctly identifies the classification and where it does not. The goal is to maximize the correct classification (green boxes) whilst minimizing the incorrect classifications (red boxes) until the pre-defined accuracies have been met.

		<u>Actual</u>	
		Accept	Reject
<u>Predicted</u>	Accept	2371	86
	Reject	61	1725

Figure 4 – Confusion Matrix, first iteration.

This figure illustrates the results of the machine learning model after only a single iteration. The model performed relatively well in correctly classifying most connections. However, there was a relatively high number of incorrect

classifications. More importantly, there was a higher number of rejectable graphs incorrectly classified as acceptable than the initial criteria allow. As previously stated, the model should err towards the conservative to avoid potential well integrity issues. Subsequent iterations were performed to eliminate these inaccuracies from the system and to optimize performance to acceptable standards.

Deployment

Once the trained model has been validated, it can be deployed to enable its use. Deployment may take on many forms ranging from incorporation as a sub-routine within existing software to creation of entirely new enterprise-level applications centered on the machine learning engine. And with the bulk of computing horsepower being required during the data processing and training stages, deployment to less powerful devices such as smartphones and personal computers is an option. Regardless of the manner of deployment, it is important to have in place processes and mechanisms to assess performance of the model, and when necessary, make adjustments for continuous improvement.

The machine learning system described in the current example for automated connection integrity evaluation was successfully deployed by Frank's International as the Intelligent Connection Analyzed Make-up (iCAM™) technology (*Figure 5*). Deployment occurred after refinement, fine tuning, and optimization of the basic system described herein. The first field deployments subjected the technology to a wide range of input conditions, all of which were handled successfully by the software with higher accuracy and reliability than human operators. The performance of this technology has proven that insights gained through data analytics and automation through machine learning can provide value-added improvements to well integrity and process efficiency.

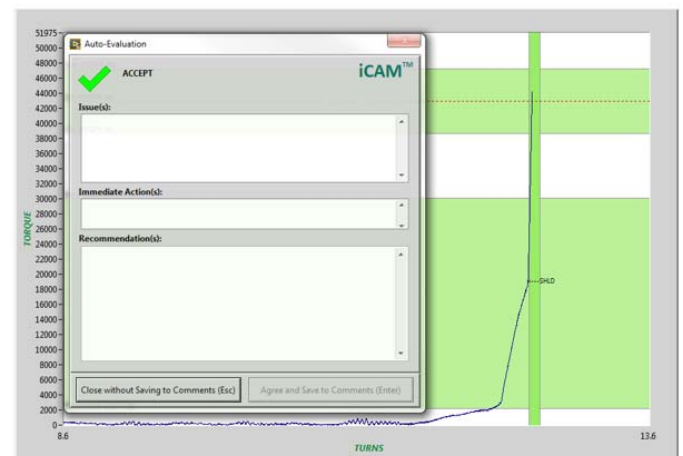


Figure 5 – Intelligent Connection Analyzed Make-up (iCAM™)

Conclusions

This paper provides an example of a real-world data

analytics and machine learning project within the Oil and Gas casing and tubular running services industry. Existing connection data were analyzed to test hypotheses, identify trends, and provide feedback of operational performance. These data were further utilized in the construction of a machine learning model capable of automatically evaluating and assessing connection integrity. Upon completion, this technology was successfully deployed to the field; realizing significant increases in reliability, integrity, and efficiency as compared to human operators. Because of these and similar advantages, the roles of data analytics and machine learning within the Oil and Gas industry are expected to increase significantly year after year.

Acknowledgments

The authors wish to thank Frank's International for its permission and support to publish this paper.

Nomenclature

ML = Machine Learning

AI = Artificial Intelligence

CAM = Connection Analyzed Make-up

iCAM = Intelligent Connection Analyzed Make-up

SME = Subject Matter Expert

RIH = Run-In-Hole

TRS = Tubular Running Service

References

1. B. Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read", in *Forbes*, May 21, 2018.
2. M. Savini, "Big Data, Machine Learning, and Well Integrity: Applications to Real World Problems," in *Engenious Symposium & Exhibition for Upstream Innovation*, 2018.
3. B. Domec, "Machine learning streamlines tubular connection analysis," in *Oil & Gas Engineering*, November 27, 2018.