# *Prediction of Rate of Penetration in Drilling*

*Abir lal Roy*

# *Prediction of Rate of Penetration in Drilling*

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology
in
Cryptology and Security

by


**Abir lal Roy**
[ Roll No: CrS 1910 ]


under the guidance of


**Yogalakshmi Sekar**
Senior Data Scientist
iLink Systems

**Dr. Anisur Rahaman Molla**
Assistant Professor
R C Bose Center for Cryptology and Security



**Indian Statistical Institute**
**Kolkata-700108, India**

**July 2021**

*With the hope of a COVID free world*

# CERTIFICATE

This is to certify that the dissertation entitled **"Prediction of ROP in Drilling"** submitted by **Abir lal Roy** to Indian Statistical Institute, Kolkata, in partial fulfillment for the award of the degree of **Master of Technology in Cryptology and Security** is a bonafide record of work carried out by him under our supervision and guidance. The dissertation has fulfilled all the requirements as per the regulations of this institute and, in our opinion, has reached the standard needed for submission.


**Yogalakshmi Sekar**
Senior Data Scientist,
iLink Systems,
ORIGIN TOWER, 11 – 13, Type 2, Dr. V.S.I Estate, Rajiv Gandhi Salai,
Thiruvanmiyur, Chennai – 60004

**Anisur Rahaman Molla**
Assistant Professor,
R C Bose Centre for Cryptology and Security,
Indian Statistical Institute,
Kolkata-700108

# Acknowledgments

I would like to show my highest gratitude to my advisors, *Yogalakshmi Sekar*, Senior Data Scientist, iLink Systems, and *Prof Anisur Rahaman Molla*, Cryptology and Security Research Unit, Indian Statistical Institute, Kolkata, India for their guidance and continuous support and encouragement.

I would also like to thank *Akhita Babu and Abhishek Soni*, consultant at iLink Systems for their valuable suggestions and discussions.They have made my works a lot easier to me with their important suggestions.

My deepest thanks to *Niladri Banerjee* of Indian Statistical Institute, Kolkata for his valuable suggestions and discussions throughout my work.

Finally, I am very much thankful to my parents and family for their everlasting supports.

**Abir lal Roy**
Indian Statistical Institute
Kolkata - 700108 , India.

# Abstract

Drilling has become an expensive and necessary operation to explore petroleum and natural gases.The goal is to increase drilling speed with minimum cost and at the same time maintaining safety.Predicting rate of penetration (ROP) in real time is very important to optimise drilling cost , since the greater the ROP is, the lesser the drilling cost would be.In this work,the typical extreme learning machine (ELM) and an efficient learning model (Artificial Neural Network) have been used in ROP prediction.Since the relationship between ROP and the set of parameters affect ROP is highly non linear, these learning models have been used to capture the non linearity.

The models have been built using WITSML Realtime drilling data which is open source data and hence erroneous.Results indicate that both ELM and ANN are competent for ROP prediction, though ELM has higher learning speed compared to ANN.Though the quality of the data set is not upto the mark, still we managed to achieve around 70% and 62% MAPE (Mean absolute percentage error) score for ELM and ANN model respectively.This work will help drilling engineers to predict ROP according to their computation and accuracy demand.

**Now we shall discuss about a different work on computer vision.**

Sewer pipeline networks have become the main concern of modern municipalities around the world as these networks are too old and they are reaching their design lifetime; meanwhile, increasing environmental and health requirements, growing demands, and tight budgets have all made the problem harder to deal with. In order to prevent severity and costly damage, sewer system conditions need to be monitored through a timely and comprehensive periodic assessment. Currently Manual inspection is common practice in the inspection and assessment of sewer networks. Visual inspection requires hundreds of hours of data processing by certified inspectors to detect defects (i.e., crack, joint offset, roots, deposit, infiltration, etc.) and assess defect severity (i.e., length, number, consequences, etc.). However, manual inspection used in the assessment of extensive sewer systems is error-prone, subjective, and time-consuming. And due to this sometimes-different type of defects in sewer pipe system go undetected until they do a good bit of damage. Here our objective is to automation model development using computer vision techniques for sewer condition assessment and automation of classification and detection of different type of defects (i.e., Crack, Fracture, Obstruction etc) which are found in sewer pipe system.

# Contents

# Chapter 1

# Introduction

## 1.1 Background

An oil well is a boring in the Earth that is designed to bring petroleum oil hydrocarbons to the surface. Wells are created by drilling down into an oil or gas reserve that is then mounted with an extraction device such as a pumpjack which allows extraction from the reserve. Creating the wells can be an expensive process, costing at least 100s of thousands of dollars, and costing much more when in hard to reach areas, i.e. when creating offshore oil platforms. The process of modern drilling for wells was made more efficient with advances to oil drilling rigs.

In the drilling industry, the rate of penetration (ROP), also known as penetration rate or drill rate, is the speed at which a drill bit breaks the rock under it to deepen the borehole. It is normally measured in feet per minute or meters per hour, but sometimes it is expressed in minutes per foot.

The industry's previously perceived ROP effect on drilling efficiency is such that if ROP increases, efficiency increases. Hence the parameters ROP is an important performance qualifiers (PQ) in this data. Hence our target in this task.

## 1.2 Our Contribution

Our contributions are summarized as follows.

- We have built two different models for prediction of ROP in drilling.The data set used to train the model is open source Volve Oilfield Wells data which is stored in WITSML format.It required a lot of effort to understand the data set and set of parameters present in the data set.There are many parameters which are irrelevant for ROP prediction.Moreover, since it is a open source data, percentage of missing value is high for most of the parameters.So it is important to choose the right well log data for training of the models.In this regard a pivot table was made to decide which well log data we should go with.In our work ,the target data set is the combination of multiple well log files each having 10,000 rows in it.

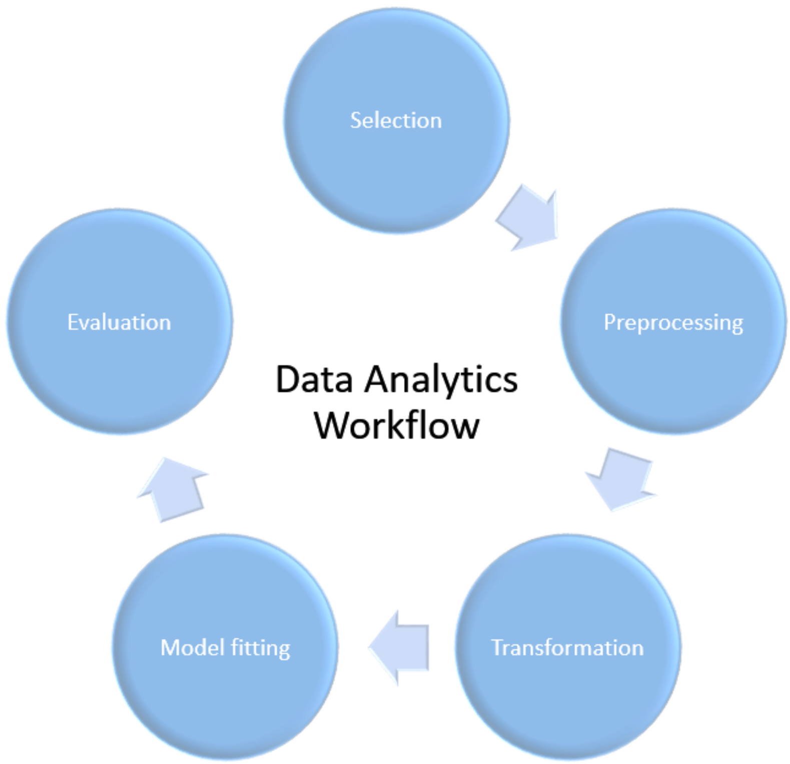



Figure 1.1: Data Analytics workflow

# Chapter 2

# Preliminaries

In this section we will discuss the preliminary settings and mathematical assumptions which we will need to describe our work.

## 2.1 Artificial Neural Network (ANN)

- An artificial neural network (ANN) is the component of artificial intelligence that is meant to simulate the functioning of a human brain.
- Processing units make up ANNs, which in turn consist of inputs and outputs. The inputs are what the ANN learns from to produce the desired output.
- Back-propagation is the set of learning rules used to guide artificial neural networks.

## 2.2 Extreme Learning Machine (ELM)

- A fast algorithm for single hidden-layer feedforward neural networks (SLFNS).
- The way ELM trains SLFN is that it first randomly generates the weights of the hidden layer and then calculates the weights of the output layer by solving a linear system using the least square method.

### 2.2.1 Algorithm

Let input vectors be $X$, dimension of the input vector be $D$, total sum of training samples be $N$, number of hidden units be $L$, dimension of the output vector be $C$ and output be $y$. Then

$$y_i = U^T h_i$$

where $U$ is $L$x$C$ weight matrix at the upper layer and $h$ is hidden-layer output

$$h_i = \sigma(W^T x_i)$$

where $W$ is $D$x$L$ weight matrix at the lower layer, T is the target vector and the kernel function is given by $\sigma(.)$

While training the model, parameters $U$ and $W$ are learned to minimize the square error $E$ given by the formula

$$E = ||Y - T||^2 = Tr[(Y - T)(Y - T)^T]$$

where the upper-layer weights $U$ can be determined by setting the gradient

$$\frac{\partial E}{\partial U} = \frac{\partial Tr[(U^T H - T)(U^T H - T)^T]}{\partial U} = 2H(U^T H - T)^T$$

to zero, leading to the closed-form solution

$$U = (HH^T)^{-1} HT^T$$

But there is an ill-conditioned hidden layer matrix $H$ (i.e., is singular). To tackle this, according to ridge regression theory, to control the degree of regularization, add a positive value $I/\mu$ to the diagonal of $HH^T$ where $I$ is identity matrix and $\mu$ is a positive constant.

Hence,

$$U = (I/\mu + HH^T)^{-1} HT^T$$

Thus,

$$minimizes \qquad ||U^T H - T||^2 + \mu ||U||^2$$

where $\mu||U||^2$ is the L2 regularization term.

Thus, y more stable and tends to have better generalization performance. The training process can thus be reduced to a pseudo-inverse problem and hence it is extremely efficient. ELM is its inefficiency in using the model parameters. To achieve good classification accuracy, ELM requires a huge number of hidden units. This inevitably increases the model size and the test time. *Note: In practice, the test time is much more valuable than the training time.*

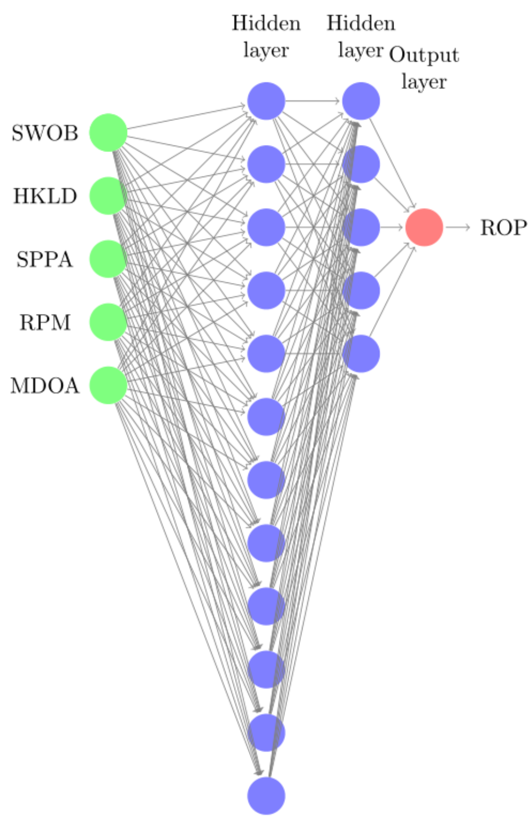



Figure 2.1: Architecture of the model

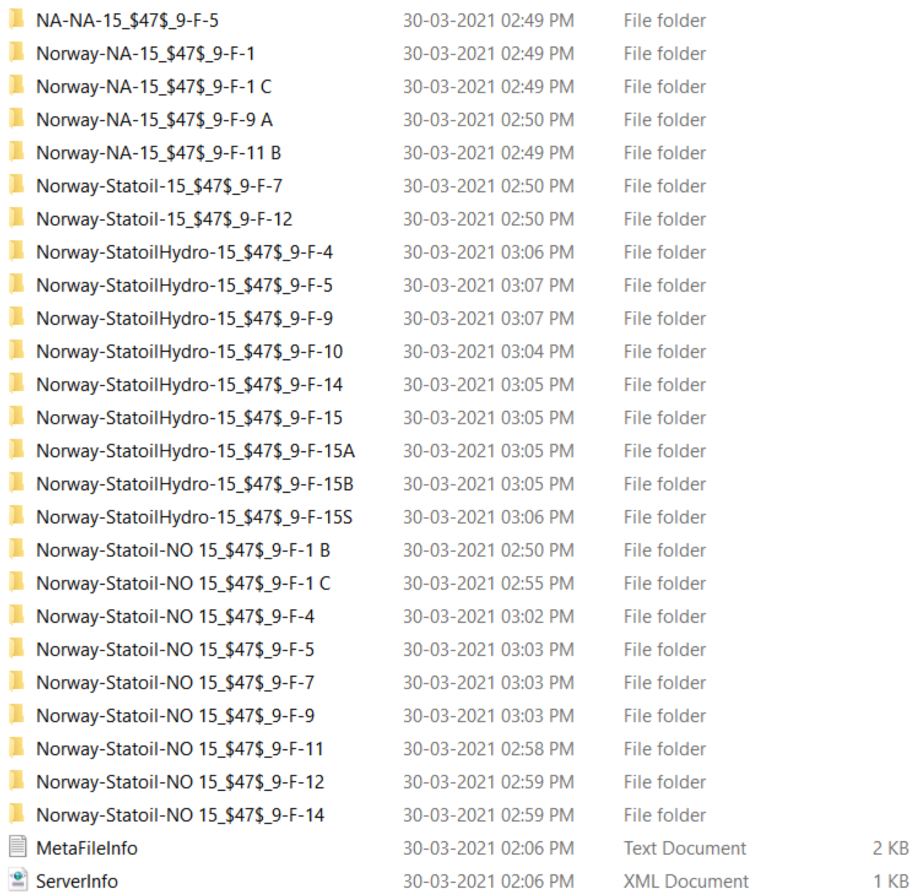
# Chapter 3

# Method for data cleaning

## 3.1 Understanding dataset

This time we will be exploring drilling data that is stored using the industry standard WITSML. This format is widely used in the industry in drilling, completion and intervention operations, specifically for real-time surveillance. WITSML stands for Wellsite Information Transfer Standard Markup Language. It has a series of rules to save the data as a consistent schema but essentially is XML. Software developers and IT professionals in the oil industry know it very well.
Since we are not writing for software developers or IT techs but for petroleum engineers, we need to know at least the basics of WITSML. We don't need to be experts on the subject but at least get familiar with the shape of the data, manipulation, perform some data exploration, and ultimately, be prepared, or know, what they are talking about when it is time to build an artificial intelligence agent based on a machine learning algorithm.
We were very fortunate of getting the Volve dataset by Equinor. It covers a wide array of data from different disciplines (production, reservoir, geophysics, drilling, completion, logging, etc.), with data coming in different formats. It is a very good exercise of data science for exploring the data and making discoveries.The dataset of the Volve Oilfield encompasses the following information :

- **Borehole Assembly Run:** The bhaRun object is used to capture information about one run of the drillstring into and out of the hole. The drillstring configuration is described in the tubular object. That is, one configuration may be used for many runs.
- **Well logs:** The log object is used to capture the curves on a well log.
- **Message:**The message object is used to provide a central location for informative time stamped information about another well related object. These messages can represent alarms, warnings, events, etc.
- **Rig:** The rig object is used to capture information about a drilling rig used to drill a wellbore.
- **Trajectory:**The trajectory object is used to capture information about a directional survey in a wellbore. It contains many trajectory stations to capture the information about individual survey points.
- **Tubular:** The tubular object is used to capture information about the configuration of a drill string.
- **Wellbore Geometry:** The wbGeometry object is used to capture information about the configuration of the permanently installed components in a wellbore. It does not define the transient drilling strings or the hanging production components.



| | | | |
|---|---|---|---|
| NA-NA-15_$47$_9-F-5 | 30-03-2021 02:49 PM | File folder | |
| Norway-NA-15_$47$_9-F-1 | 30-03-2021 02:49 PM | File folder | |
| Norway-NA-15_$47$_9-F-1 C | 30-03-2021 02:49 PM | File folder | |
| Norway-NA-15_$47$_9-F-9 A | 30-03-2021 02:50 PM | File folder | |
| Norway-NA-15_$47$_9-F-11 B | 30-03-2021 02:49 PM | File folder | |
| Norway-Statoil-15_$47$_9-F-7 | 30-03-2021 02:50 PM | File folder | |
| Norway-Statoil-15_$47$_9-F-12 | 30-03-2021 02:50 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-4 | 30-03-2021 03:06 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-5 | 30-03-2021 03:07 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-9 | 30-03-2021 03:07 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-10 | 30-03-2021 03:04 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-14 | 30-03-2021 03:05 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-15 | 30-03-2021 03:05 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-15A | 30-03-2021 03:05 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-15B | 30-03-2021 03:05 PM | File folder | |
| Norway-StatoilHydro-15_$47$_9-F-15S | 30-03-2021 03:06 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-1 B | 30-03-2021 02:50 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-1 C | 30-03-2021 02:55 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-4 | 30-03-2021 03:02 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-5 | 30-03-2021 03:03 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-7 | 30-03-2021 03:03 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-9 | 30-03-2021 03:03 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-11 | 30-03-2021 02:58 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-12 | 30-03-2021 02:59 PM | File folder | |
| Norway-Statoil-NO 15_$47$_9-F-14 | 30-03-2021 02:59 PM | File folder | |
| MetaFileInfo | 30-03-2021 02:06 PM | Text Document | 2 KB |
| ServerInfo | 30-03-2021 02:06 PM | XML Document | 1 KB |

Figure 3.1: View of all the wells that are storing data in WITSML format

## 3.2 Understanding XML format

**Extensible Markup Language** (XML) is a simple text-based format for representing structured information: documents, data, configuration. XML documents create a hierarchical structure looks like a tree.

The first line in the picture below is the XML declaration. It defines the XML version (1.4.1.1) and The next line describes the root element of the document (like trajectory, log, wbGeometry, etc) and the next lines describe the child elements of the root.

```
−<logs version="1.4.1.1">
  −<log uidWell="dd19bf7b-02a7-4383-9038-ce201cee4d91" uidWellbore="9ab4989d-11c6-4ac1-83c1-e74362292dd8" uid="11RJGJ56">
      <nameWell>NO 15/9-F-15</nameWell>
      <nameWellbore>NO 15/9-F-15 C</nameWellbore>
      <name>ReamData</name>
      <objectGrowing>false</objectGrowing>
      <serviceCompany>Baker Hughes</serviceCompany>
      <runNumber>0</runNumber>
      <pass>0</pass>
      <creationDate>2013-10-24T00:07:51.423Z</creationDate>
      <description>Data Log</description>
      <indexType>measured depth</indexType>
      <startIndex uom="m">784.8</startIndex>
      <endIndex uom="m">2637.3</endIndex>
      <direction>increasing</direction>
      <indexCurve>Depth</indexCurve>
      <nullValue>-999.25</nullValue>
      <priv_dTimPriority>2013-10-24T00:07:51.423Z</priv_dTimPriority>
    −<logCurveInfo uid="MWIN">
        <mnemonic>MWIN</mnemonic>
        <classWitsml>wtMudInAv</classWitsml>
        <unit>kg/m3</unit>
        <mnemAlias>MWIN</mnemAlias>
        <nullValue>-999.25</nullValue>
        <minIndex uom="m">784.8</minIndex>
        <maxIndex uom="m">2637.3</maxIndex>
        <curveDescription>Mud Weight In</curveDescription>
        <typeLogData>double</typeLogData>
      </logCurveInfo>
    −<logCurveInfo uid="Depth">
        <mnemonic>Depth</mnemonic>
        <unit>m</unit>
        <mnemAlias>Depth</mnemAlias>
        <nullValue>-999.25</nullValue>
```


Figure 3.2: XML format of the WITSML datatype

## 3.3 Generic Code to access WITSML data :

Beautiful Soup is a Python library for pulling data out of HTML and XML files. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree.

## 3.4 Parameter's Definition

Our work concentrates on ROP estimation using Weight on bits, Hookload, Standard pipe pressure ,Mud density and RPM .There are more important parameters which affect ROP ,but we did not take that into consideration because of their unavailability.Now we shall try to understand what these parameters actually are.

- **Weight on Bits:** Weight on Bit is the weight or the pressure you apply to cut. Think of it as compared to cutting a tomato vs cutting a pineapple. You will apply more pressure/ weight on the knife to cut a pineapple. Less weight on bit, and you will not cut anything, more weight on bit can lead to floundering, i.e. the bit spinning with too much dust. There is an optimum range to keep.

- **Standard pipe pressure** This is a very important feature. We pump drilling fluid down the drilling string and up the annulus. The pressure to pump this fluid is generated by a pump on the rig. Think of this fluid as the blood of the well and hence this pressure is the blood pressure. High is bad, low is bad.

- **Mud Density:** It is mass per unit volume of drilling Fluid.

- **Hookload:** This is the value of the weight of the entire drillnig string. Hookload tells you a lot of things but it can get very complicated to read when your well is deviated. We can discuss this over a call if you want.
- **RPM:** It is Revolution per minute. It is the speed at which we rotate the drilling bit. We also use downhole hydraulic motors to achieve higher RPMS. We can spin from the surface also. This helps in having a relative motion between the bit and the earth. Think of it as the front and back motion you do with your hand while cutting an apple.
- **ROP:**In the drilling industry, the rate of penetration (ROP), also known as penetration rate or drill rate, is the speed at which a drill bit breaks the rock under it to deepen the borehole.

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | TIME - s | SWOB - kk | HKLD - kkg | SPPA - kPa | RPM - rpm | Bit_RPM - | MDOA - g/ | ROP - m/h | log | well_name | folder |
| 2 | 2007-12-1 | 0 | 62.08032 | 197 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 3 | 2007-12-1 | 0 | 62.05993 | 195 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 4 | 2007-12-1 | 0 | 62.1415 | 206 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 5 | 2007-12-1 | 0 | 62.26387 | 187 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 6 | 2007-12-1 | 0 | 62.29446 | 186 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 7 | 2007-12-1 | 0 | 62.1415 | 185 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 8 | 2007-12-1 | 0 | 62.19249 | 187 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 9 | 2007-12-1 | 0 | 62.31486 | 163 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 10 | 2007-12-1 | 0 | 62.19249 | 188 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 11 | 2007-12-1 | 0 | 62.23328 | 190 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 12 | 2007-12-1 | 0 | 62.37604 | 186 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 13 | 2007-12-1 | 0 | 62.24348 | 196 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 14 | 2007-12-1 | 0 | 62.34545 | 186 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 15 | 2007-12-1 | 0 | 62.32505 | 207 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 16 | 2007-12-1 | 0 | 62.28427 | 206 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 17 | 2007-12-1 | 0 | 62.31486 | 210 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 18 | 2007-12-1 | 0 | 62.42703 | 181 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 19 | 2007-12-1 | 0 | 62.5086 | 182 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |
| 20 | 2007-12-1 | 0 | 62.31486 | 194 | 0 | | 1.02 | 53.39 | 1 | NA-NA-15_ | 1 log 1 2 1 |

Figure 3.3: View of well log data

## 3.5 Data Wrangling

Data wrangling defines the steps that are required to transform raw data (numeric or textual) into a format that it can be readily used for analysis, modelling, statistics, etc.Now we will describe the data wrangling steps needed to clean drilling data.

### 3.5.1 Cleaning Missing value

One of the major challenges in most data science projects is data quality (or lack thereof). In fact, most project teams spend 60 to 80 percent of total project time cleaning their data.After combining multiple log files together ,we had very less amount of missing data compared to size of the data set.That's why it has been discarded from the data set.

| | Parameter | Data Type | Sum | No of Uniques | Missing Count | % missing |
|---|---|---|---|---|---|---|
| 0 | TIME - s | object | Not Applicable | 59487 | 0 | 0.000000 |
| 1 | SWOB - kkgf | float64 | 319743 | 1723 | 8 | 0.006724 |
| 2 | HKLD - kkgf | float64 | 1.04047e+07 | 5469 | 8 | 0.006724 |
| 3 | SPPA - kPa | float64 | 7.72199e+08 | 10163 | 8 | 0.006724 |
| 4 | RPM - rpm | float64 | 7.72687e+06 | 394 | 8 | 0.006724 |
| 5 | Bit_RPM - rpm | float64 | 1.45137e+06 | 106 | 107108 | 90.018826 |
| 6 | MDOA - g/cm3 | float64 | 120726 | 2 | 8 | 0.006724 |
| 7 | ROP - m/h | float64 | 4.92018e+06 | 1788 | 8 | 0.006724 |
| 8 | log | int64 | 770856 | 12 | 0 | 0.000000 |
| 9 | well_name | object | Not Applicable | 1 | 0 | 0.000000 |
| 10 | folder | object | Not Applicable | 1 | 0 | 0.000000 |

### 3.5.2 Outlier Treatment

Outlier is commonly used terminology by data scientists.An outlier is a data point in a dataset that is distant enough from all other observations.As expected here in our data ,outliers were present.The outliers were detected using inter quartile range which is defined as **IQR** $= Q_3 - Q_1$ where $Q_3$ is the 3rd quartile and $Q_1$ is the first quartile.Following the definition , outliers are defined as any data points that are outside the range of 1.5 times the interquartile range above the 3rd quartile and below the first quartile.In our case 75 percentile and 25 percentile were used for removing outliers.

```
HKLD - kkgf      6914
MDOA - g/cm3        0
ROP - m/h        3368
RPM - rpm           0
SPPA - kPa          0
SWOB - kkgf      2808
```

Figure 3.4: Parameter wise outlier count

## 3.6 Analysis

Literature reviews show that the relationship between ROP and set of parameters affecting ROP is highly non linear and complex.In our correlation heatmap also ,ROP is found to be very less correlated with other parameters.ROP related parameters can be classified into 3 classes viz Rig and bit related parameters ,Formation parameters and Drilling fluid properties.

| Rig and bit related parameters | Formation parameters | Drilling fluid properties |
|---|---|---|
| Weight on bit (WOB) | Local stresses | Mud weight |
| Torque | Hardness | Viscosity |
| Rotary speed (RPM) | Mineralogy | Filtrate loss |
| Flow rates | Porosity and permeability | Solid content |
| Pump stroke speed (SPM) | Formation abrasiveness | Gel strength |
| Pump pressure | Drillability | Mud pH |
| Hook load | Depth | Yield point |
| Bit wear | Temperature | |
| Type of the bit | Unconfined compressive strength (UCS) | |

Figure 3.5: ROP related drilling parameters classification

The following correlation heatmap is showing how less correlated ROP is with the other parameters and hence some soft computing techniques like Neural Network ,ELM have been used to predict ROP in order to capture non linearity.

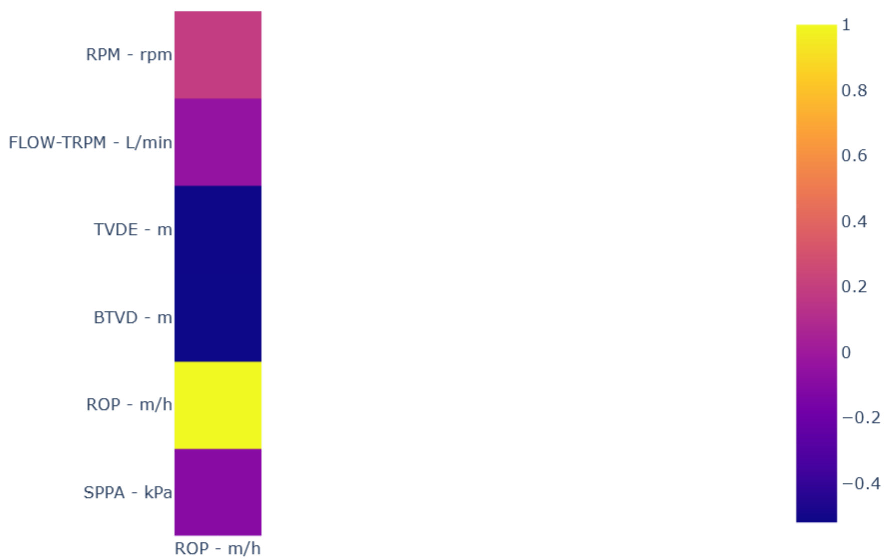



Figure 3.6: Coorelation Heatmap

The distribution check up for ROP has been also done.A smooth density plots was used to understand the probabilistic behaviour of ROP. In this plot some sharp edges are found at interval

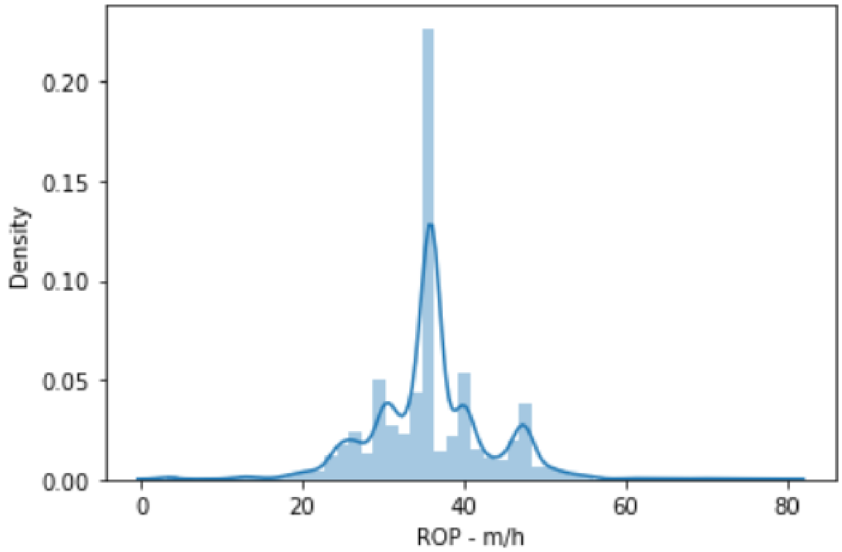



Figure 3.7: Distribution of ROP

boundaries.Transformation technique such as square root,log transformation ,Box cox transformation etc can be used to remove this peak and to make the data follow normal distribution.

Now we shall see the variation of ROP with respect to time.In the graph below,it is found initially that ROP is 0 at the beginning and as the time goes ,variations are found with respect to time.In the graph different colors are used to indicate that the data belongs to different well log.It takes around 2-3 months to complete the entire drilling procedure for a single well.Each well-log xml file contains data of around 8 hours of drilling.We have analysed 5 well-log data sets belonging to single well.

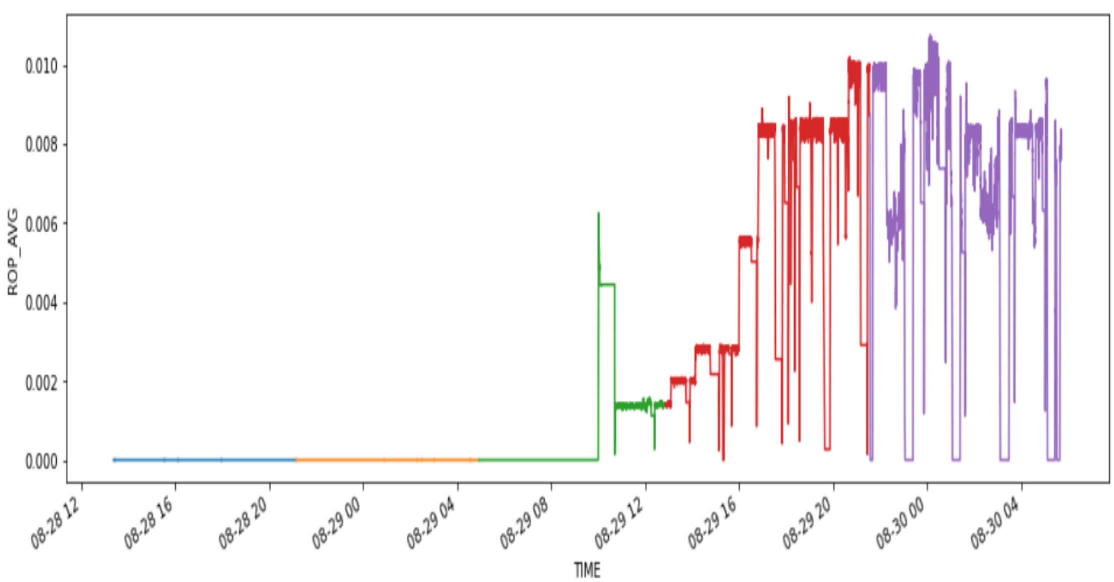



Figure 3.8: Time Vs ROP

# Chapter 4

# Conclusion

## 4.1 Results

This study scrutinizes the feasibility of ANN and ELM using the recorded data of the reservoir in the Volve field to predict the ROP of offshore drilling operations. This ANN and ELM models can successfully predict ROP in real time, and this prediction can be used as a reference range for drilling engineers identifying drilling problems and making decisions.

The figure 3.1 depicts the performance of ANN and the figure 3.2 depicts the performance of ELM on the same data.

The results of the ELM model were compared to those of the classical ANN model. Performance of the models was checked by using Mean Squared Error (MSE) and Mean Absolute Error (MAE) performance indicators. According to the performance indicators, all of the neural networks are competent for ROP prediction, but the prediction performances of the ELM model were found to be better than those of the other ANN model with respect to accuracy and also had the lowest running speed. Compared to conventional ANN models, the result of this work shows the potential of some fast and flexible neural network approaches to model the ROP with high accuracy while maintaining running speed. Therefore, drilling engineers can make better choices according to accuracy and computational demand in practical use.

## 4.2 Discussion

Most companies already have the data to perform data analysis in drilling. The accessibility and usage of unstructured data for regulatory, analytic and decision-making purposes is driving the need to search and scrutinize this data. An organized workflow to treat the massive data influx, and a systematic way to evaluate previous performance against well defined metrics is required. Data preparation which include data pre-processing, profiling, cleansing, validation and transformation is the most difficult and time consuming task in the whole workflow. A detailed workflow to guide a drilling engineer to clean drilling data is presented in this work that can result in significant time savings from the raw data collection to data analysis. Once the drilling engineer has the data cleaned and structured, a vast number of analysis can be performed with the data to extract valuable information about previous drilling performance. The close monitoring of actual and previous operational parameters can result in significant performance gains.

Table 4.1: Performance comparison of ANN and ELM

| Models | Accuracy |
|---|---|
| Classical Artificial Neural Network | 61.04 |
| Extreme Learning Machine | 68.80 |

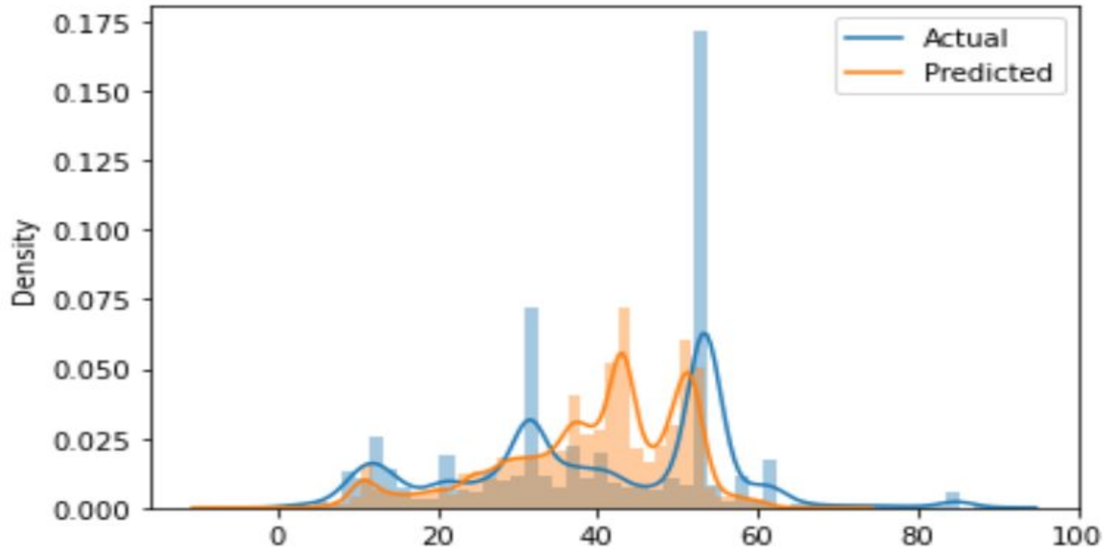



Figure 4.1: Performance of ANN

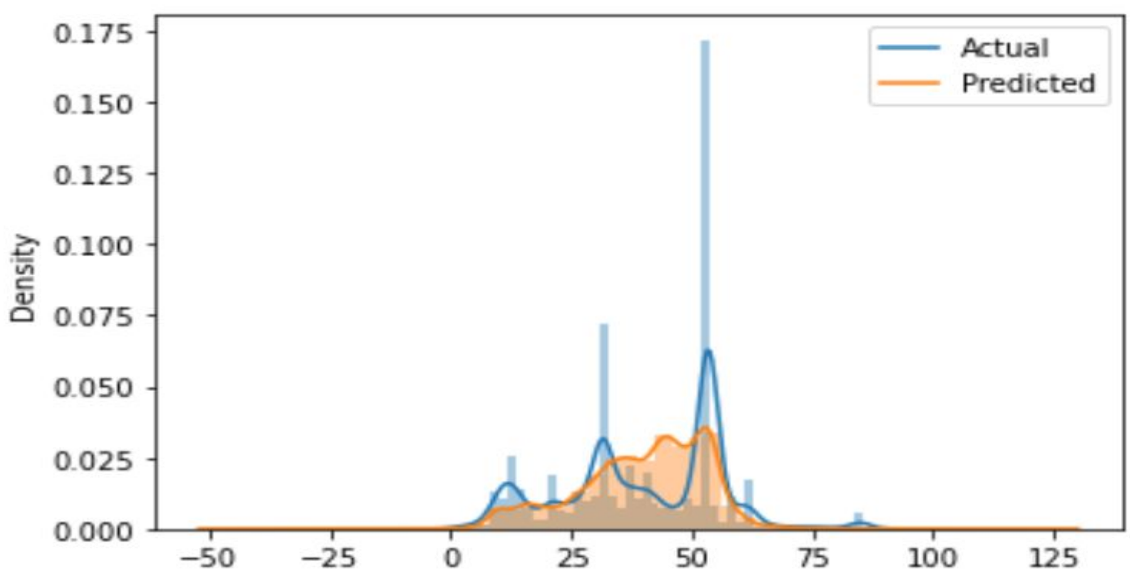



Figure 4.2: Performance of ELM

Choosing the relevant parameters is typically difficult. But parameters like Weight on Bit (WOB) , Hook Load, Flow Rate, Pump Pressure, Revolutions per Minute (RPM), Torque, Depth and Mud Density were selected as relevant parameters that affect ROP and were also involved in developing neural network models. However, for systems with data values beyond the range of this study or for a new bit, it is necessary to develop new neural networks. Conversely, the wider range of input can enable the neural network models developed to have wider applications in select cases.

## 4.3 Future Scope

In recent years, the offshore oil and gas industry has changed in eager haste, with new technologies being adopted by the energy sector to meet the challenges of a digital economic landscape. Artificial intelligence is an exciting new technological field, explores its applications in the offshore oil and gas industry. This refines the process through iterations to produce programs tailored to specific purposes that allows companies to monitor complex internal operations and respond quickly to concerns that human operators may not have been able to detect.

# Chapter 5

# Object Detection using Yolo model

## 5.1 Problem Statement

Closed Circuit Television is very commonly used technology to inspect the sewer pipe condition. To find the problem running in a sewer pipe we use camera (CCTV to detect the weak point and fix it later. It is very obvious that we should make proper diagnose in order to prevent unnecessary repairing and reinstallation of the sewer pipe. But after going through the video, manual inspection of defect is very time consuming, labor intensive and tend to have error in it. To assist the inspector and to save a significant amount of time we are building an object detection model which will automatically classify the defects with high accuracy score.

## 5.2 Available data and data Distribution

Till now we have worked on around 500 videos. First, we have extracted the images of the defects from the videos also we recorded the different defect occurrence of each video. From that we found enough defect occurrences for 28 defect classes.

- From the available data for the 28 defect classes, we were able generate 7000 Images (250 images per defect).
- Then image annotation step takes place to map all the image with their respective defect classes.

| Category | Defect Class |
|---|---|
| Broken | B |
| Cracks | CL , CM , CS |
| Deposits | DAE, DAGS, DAR, DSC, DSGV, DSF |
| Fractures | FL, FM, FS |
| Hole | HSV, HVV |
| Obstruction | OBB, OBI, OBR |
| Roots | RMJ |
| Taps | TB, TBA, TBC, TBD, TBI, TFA, TFC, TFD, TRB |

## 5.3 Approach and Architecture

Object detection is a task in computer vision that involves identifying the presence, location, and type of one or more objects in a given photograph. We are using Yolo Algorithm to deal with this problem.

- **<u>Yolo Model</u>**: **YOLO ("You Only Look Once")** is an effective real-time object recognition algorithm which provides the tools for doing just that – finding all the objects in an image and drawing the bounding boxes around them.YOLO trains on full images and directly optimizes detection performance .

- **<u>Advantages of Yolo</u>**: This model has several benefits over other object detection methods:

  1. YOLO is extremely fast (Speed: 45 frames per second — better than real-time)
  2. YOLO sees the entire image during training and test time, so it implicitly encodes contextual information about classes as well as their appearance.

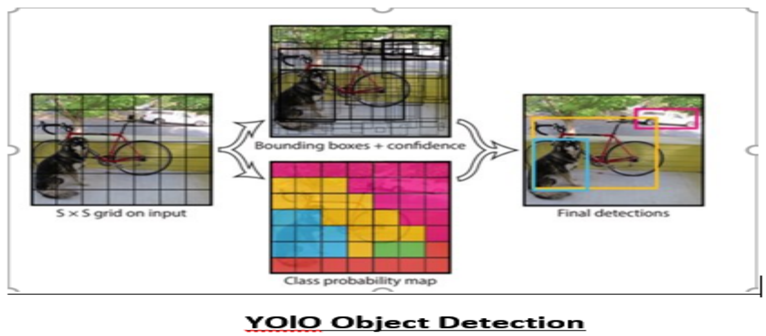



Figure 5.1: **Source :** Google

- **<u>Convolution Neural Network (CNN)</u>**:CNN is a Deep Learning algorithm where the input is an image and convolutions with various kernels occur in each layer to produce the convolved image as the input into the next layers. There are often other types of layers in better convolutional layers such as pooling layers that extract the important information/features from the previous layer's convolutions.The Steps Involved in the CNN can broadly be classified as:

$$InputImage \rightarrow Filtering \rightarrow Relu \rightarrow Pooling \rightarrow ... \rightarrow Vectorization \rightarrow FCNN$$

  Where FCNN stands for fully connected layer. For data preparation and model training

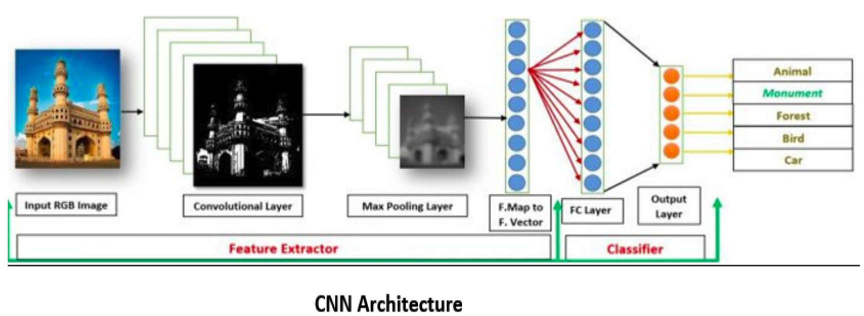


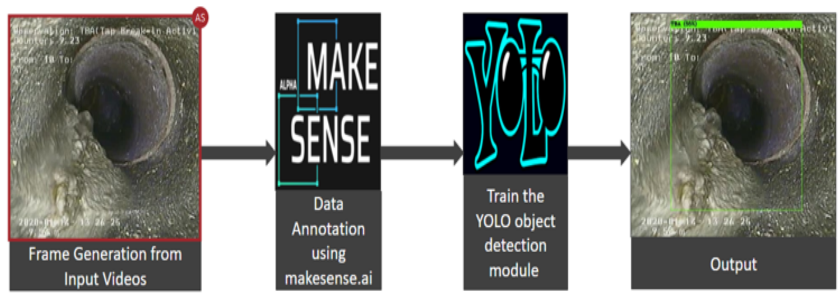

Figure 5.2: CNN Architecture, **Source :** Google

  following steps have been taken :

  – Extract the frames in which a particular defect is present from video.

- Annotate the extracted images and save it in yolo format. It returns a text file with coordinate of the bounding box.
- . Once the data is in our hand, we feed it to the model for training.. Once the data is in our hand, we feed it to the model for training.

- **Model Output**:




## 5.4 Conclusion

Till now, we have been able detect 28 different defects in a sewer pipe system with moderate confidence score. We have to explore different sampling technique to handle class imbalances without losing any information .Also we are trying to automate the process of data preparation what we did manually and increase the number of detected defect classes.

# Bibliography


[1] An Efficient Approach for Real-Time Prediction of Rate of Penetration in Offshore Drilling by Xian Shi, Gang Liu,Xiaoling Gong, Jialin Zhang, Jian Wang, and Hongning Zhang

[2] Field Drilling Data Cleaning and Preparation for Data Analytics Applications by Daniel Cardoso Braga

[3] "https://www.discovervolve.com/2021/01/06/understanding-the-witsml-format-in-the-volve-oilfield-dataset/"

[4] "http://blog.oilgainsanalytics.com/publications/exploring-drilling-data-from-volve-dataset-witsml-r/"

[5] "https://docs.opencv.org/master/"

[6] "https://pjreddie.com/darknet/yolo/"