(Thesis Proposal)
Failure prediction in open-hole wireline logging of oil and gas drilling operation using support vector machine.

Miss Maylada Pootisirakorn

A Thesis Proposal Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2018

(โครงร่างวิทยานิพนธ์)

การทำนายความผิดปกติของการหยั่งเชิงธรณีของหลุมขุดเจาะน้ำมันและแก๊สโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน

น.ส.เมลดา พุฒิศิรกร

โครงร่างวิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561

Thesis Proposal Title	Failure prediction in open-hole wireline logging of oil and gas drilling operation using support vector machine.
By	Miss Maylada Pootisirakorn
Field of Study	Computer Science
Thesis Advisor	Professor PRABHAS CHONGSTITVATANA, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

..... Dean of the Faculty of Engineering
(Associate Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS PROPOSAL COMMITTEE

..... Chairman
(Assistant Professor SUKREE SINTHUPINYO, Ph.D.)
..... Thesis Advisor
(Professor PRABHAS CHONGSTITVATANA, Ph.D.)
..... External Examiner
(Associate Professor Worasait Suwannik, Ph.D.)

TABLE OF CONTENTS

Page

1. Introduction

Oil and gas exploration and production is a process based on various kind of data that can describe the unknown surface under the ocean to drill. The operation requires data to make plan and decision. Each of operation requires high cost with operational excellence, however the offshore operation, under the sea level has high pressure, temperature which affects to the operation.

In the past the data from oil and gas field is mostly used for descriptive and diagnostic purpose, however the focus has been changed to do more on predictive and prescriptive as the machine learning can be applied to get insight from the big amount of data that already kept but have not been used by human in the past. [Figure 1]

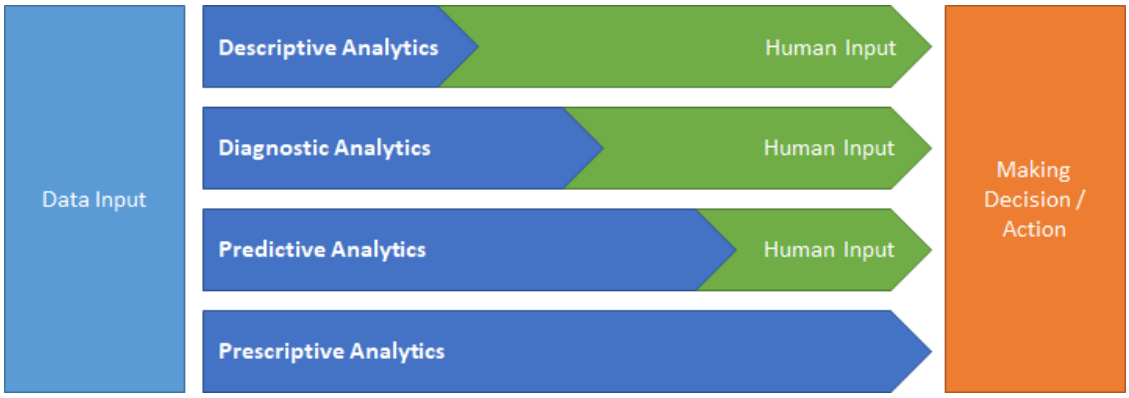

Figure 1 Data analytics focus types

The open hole wireline logging usually run after drilling to target depth with the decision of engineers to run the logging in the next step based on their experience and a standard procedure. After drilling follow the plan, open hole wireline logging will help engineers interpret the real situation from the logs. Log data has been collected from the tools, however, sometimes the tools are stuck in the hole and could not go to the expected target depth, this leads to the loss of rig time and operation time and it could be worst if the tool is lost the connection from the wire.

An unexpected cost will be added to get the tools out of the hole prior to resume to normal operation. It was found that the problems encountered it may cause from several possible issues such as temperature, pressure, directional surveys, formation, fluid density including the circulation. Engineers could not know exactly that the logging will success or failure until the result found out after sending the log tools.

Machine learning method can be used for predicting this fault based on historical data. The method is applied to the operation data that collected after completed drilling to target depth. Support Vector Machine (SVM) [1] is a popular technique in machine learning that can be used to classify two-classes data. The data set has a small rate of failure which is lead to the imbalance data set. This research is

focus to train SVM to find out the probability of the failure that might happens before running the open-hole wireline logging to save time and cost.

2. Theoretical Background

2.1 Offshore Drilling Operation

Offshore drilling operation [2] refers to the process to drill through soil and rock under the seafloor to create a well which is bored hole that can access to geological reservoirs contained with oil and gas. The development or production wells are drilled to recover oil and gas reserves in the proven economic areas.

The process of drilling oil and gas well involves several steps:

1. A well is drilled using drill bit and pipe to create a bore hole under the seafloor. The drilling path could not be drilled directly to hit the oil and gas reservoir otherwise it would be blow out or explode before doing the completion and production. It is done by boring a vertical depth with angled to the target reservoir.
2. The circulation process in the hole using mud to circulate and remove the rock cuttings from the hole and maintain the working temperature and pressures of the well.
3. Cementing requires on each section after drill to the planned depth. This is applied to the bore hole to prevent collapse. There are mainly three sections of the well in Gulf of Thailand.

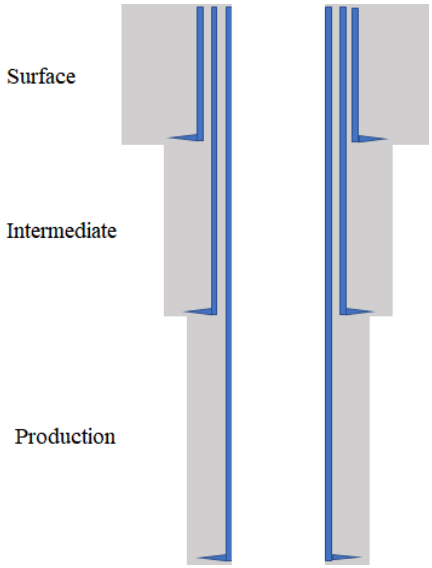

Figure 2.1 Schematic of the well with cementing on three major sections; surface, intermediate ant production.

4. Once the well is drilled to the target depth at production section on bore hole or called open hole before cementing, it usually has the open hole wireline logging

or formation test after pulling the drill bit out of the hole. Wireline logging is the process to collect data using the electric instruments to continuously measure the properties of a formation, this data can help making decisions in drilling and production operation.

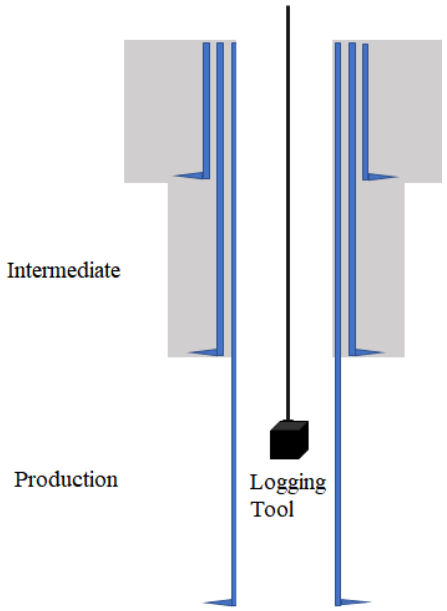

Figure 2.2 Wireline logging tool in bored hole or open hole

5. After the well has been drilled to the target depth with casing cement, it is ready for completion and production. Completion activity called perforation creates small hole through the casing. The small hole is passed through the production area that connect to the reservoir. This provide a flow path for oil and gas.

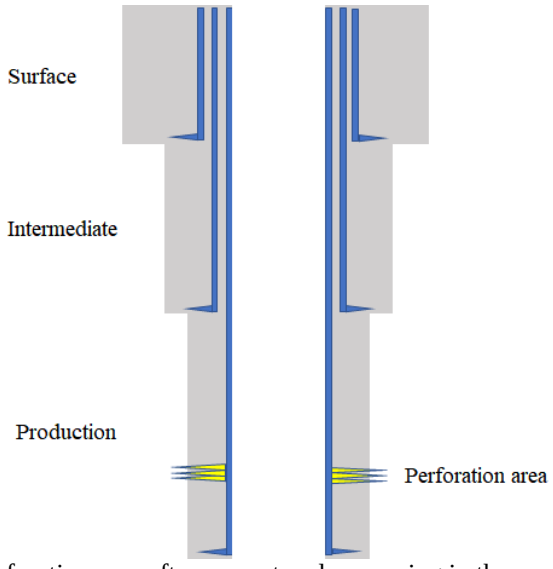

Figure 2.3 Perforation area after cement and run casing in the production section

6. Then the production process continues by maintaining the valves and pump to produce the oil and gas on the platform area.

2.2 Open-hole wireline logging

To drill the well, it is a technological process however no wells are identical with various of risk due to the temperature and pressure is increased when drilled to the deeper hole. The information of the subsurface around the hole can be acquired from the electronic logs which represent important source of data to geophysicist and engineers to analyze and explore the rock information and the reservoir target which can be produced the oil and gas.

The open hole logging activities [3] is one of a large investment that oil and gas company made to acquire the data. It is important if we can reduce the cost of this activities and ensure that it would not be failed and reduce the non-productive time of the drilling operation. The type of logging [4] to the open hole and objective of the data acquisition are the two main questions that need to be verified prior to start the logging operation since the data gathering would help expert to interpret and making consideration using statistical skill about well integrity and reservoir characteristics. However, it is not a rule for data acquisition of the logging operation of every well, the number can be reduced depends on the hole condition since the logging through casing options still exist. Data that could be get from the open hole are such as assessment of source rock potential, hole volume and shape estimates, sample of lithology, location of hydrocarbon, reservoir capacity assessment, porosity and pressure measurements.

Type of logging

1. Formation Tester is mainly used for collecting the pressure point for specific depth, the different subtype depends on temperature and the service company.
    (1) RDT (Reservoir Description Tool)
    (2) HSFT (Hostile Sequential Formation Tester)
    (3) HXPT (High Temperature Xpress Pressure Tool)
    (4) SRFT (Slimhole Repeat Formation Tester)
2. Wireline Logging is used primarily to identify lithology, reservoir porosity and fluid type in formation.
    (1) Quad Combo
    (2) Triple Combo

Quad Combo provides more detail of time-depth and pore pressure which is typically run is delineation or exploration wells to retrieve more information for making decision.

Open hole logging operation will be executed after drilled to the production section, there are 6 hours before the operation happen. Actual of parameters after drilled can be used to analyze and make decision prior to run the logging.

2.3 Support Vector Machine

Support Vector Machine (SVM) [5] is a supervised machine learning algorithm which means that the answer must be known and use for training and building a predictive model. There are two main types for SVM, one is Classification, and another is Regression which can handle multiple continuous and categorical variables. Classification is used for this research. The classification SVM construct the optimal hyperplane that separate data into groups with minimized error function through training process.

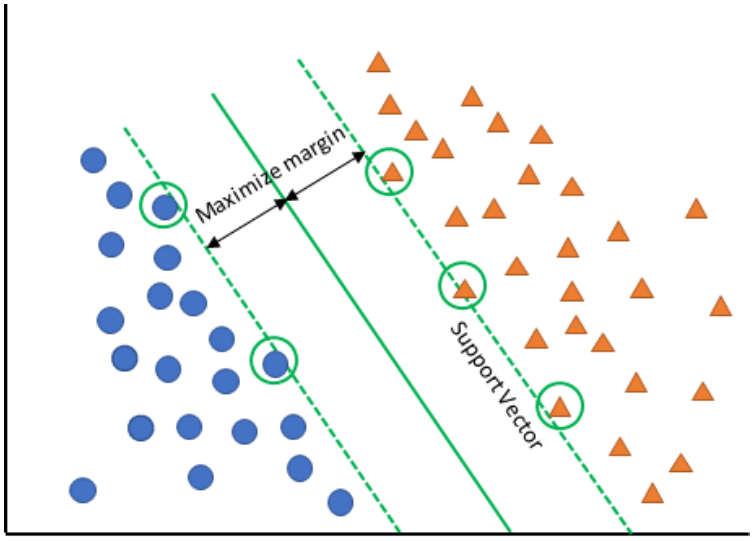

Figure 2.4 Support Vector Machine that separate the hyperplane to classify the class with maximum margin

2.4 Confusion Matrix

Confusion matrix [6] is a technique to summarize and describe the performance of a classification model on a test data set. It is use in evaluation method of the classification problem. The number of correct and incorrection predictions are summarized in table of numbers for each class. It gives an insight into the errors from the model but moreover the types of errors are also important which is not only the classification accuracy.

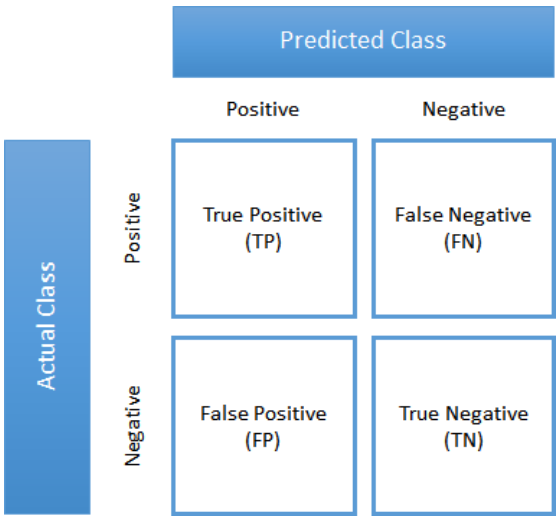

Figure 2.5 Confusion matrix as a table summary for a binary classification problem

- **Actual Class** is the result as labeled by actual experiment
- **Predicted Class** is predicted results by the test
- **True Positive (TP)** is the number of samples that were predicted as positive and the actual result is labeled as positive.
- **False Positive (FP)** is the number of samples that were predicted as positive, but the actual result is negative. It is also called Type-1 error.
- **True Negative (TN)** is the number of samples that were predicted as negative and the actual result is also negative.
- **False Negative (FN)** is the number of samples that were predicted as negative, but the actual result is positive. It is called Type-2 error.

Moreover, from number that represent in confusion matrix, there are other performance measurement that can be used to evaluate the model as below:

- **Accuracy** measures overall accuracy of the model classification

$$Accuracy = \frac{\text{all correct}}{\text{all}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}$$

- **Precision** is the positive predictive value

$$Precision = \frac{\text{True Positive}}{\text{Predicted Possitive}} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Recall** is the ratio of correct predictions and the actual positive

$$Recall = \frac{\text{True Positive}}{\text{All Possitive}} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

3. Literature Review

There have been increasing research activities related to apply machine learning techniques to predict results or detect errors in the field of oil and gas industry to get insight and help in making decision from the data during the past decade.

3.1 Machine Learning Overcomes Challenges of Selecting Locations for Infill Wells

The recent research that use machine learning to help in selecting locations for infill well in 2018 [7] using the support vector machine technique to the data collection the regularization parameters were determined using grid search to prevent overfitting. The SVM model was trained to rank the locations based on their production capabilities and historical of reservoir data and completion data, the new workflow after has been applied to help the asset team making data driven decisions.

3.2 Data Driven Approach to Failure Prediction for Electrical Submersible Pump System

In 2015 [8] presents a data driven approach for failure prediction of the pump system that used in oil and gas industry using support vector machine technique to train the selected features and test on real world data. The data collected by sensors based on electrical and frequency data and other information such as logs are feed to the framework to predict the results. The successful and timely diagnosis of failure from the model can improve the production performance. The paper selected the SVM as it is powerful binary classifier and using the feature extraction and selection in their work.

3.3 Implementing Artificial Neural Networks and Support Vector Machines in Stuck Pipe Prediction

The research of stuck pipe prediction in oil and gas industry in 2012 by Islam et al [9] focus on using artificial neural networks (ANN) and support vector machine to predict the stuck pipe before it occurs. It is one of the most costly problem. The model was designed and constructed by MATLAB built in function and library. The study classifies stick pipe incidents into two groups as stuck or non-stuck. The SVM can predict stuck pipe occurrences with accuracy over 85% and claim that SVM is more convenient than ANN since it need fewer parameter to be optimized. The model generally works well in the selected area of the operation but may not work in other areas. Previously in 2006 Siruvuri et al [10] use ANN to predict stick pipe, the reasonable outputs were accepted even the data might be incomplete or have some errors.

3.4 Classification of Petroleum Well Drilling Operations Using SVM

SVM has been used to classify petroleum well drilling operations in 2006 Adriane et al [1] present the development of a system that intends to make better user of the information collected during well drilling operation. Main idea is to use a great about of data that has not been properly used and it might provide insight using SVM for pattern recognition and develop the automatic classification system that can produce performance enhancement. This paper presents the 6 multi-class SVM and tested by the gaussian RBF, polynomial and linear functions using MATLAB. The simple linear SVM has the good generalization accuracy with correctness of 92%.

3.5 Study on Intelligent Prediction for Risk Level of Lost Circulation While Drilling Based on Machine Learning

The well problem is one of the most interesting issue that need focus on in 2018 Zejun Li et al [11] study on three typical machine learning algorithms to analyze drilling data in Iraq to predict the lost circulation issue. There are SVM, ANN and random forest. SVM and random forest have predicted 99% or wells with normal, however the data is imbalance only 55% of the lost circulation samples are correctly classified. 45% are incorrect. For ANN in the lost circulation cases 46.6% are correctly classify and 53.4% are incorrect. Compare to overall classification accuracy, the accuracy to identify lost circulation points is not ideal, partly because they are relatively sparse which is imbalance compared to the normal samples.

Table 3.1 Machine learning techniques application in oil and gas practices

Application	**ML Technique**	**Data Set**	**Researcher**	**Year**
Selection Infill Location	SVM, K-Means Clustering	Reservoir data, Oil -production rate, and completion data.	Adam Wilson	2018
Pump Failure Prediction	SVM	Electrical and frequency data from the field	Dong Guo et al	2015
Stuck Pipe Prediction	ANN, SVM	Mud logging and well information	Islam Al-Baiyat and Lloyd Heinze	2012
Classification of Well Drilling Operations	SVM	Drilling well information	Adriane B. et al	2006
Risk Level of Lost Circulation	SVM, Random Forest, ANN	Mud logging and well information	Zejun Li et al	2018

4. Research Methodology Plan

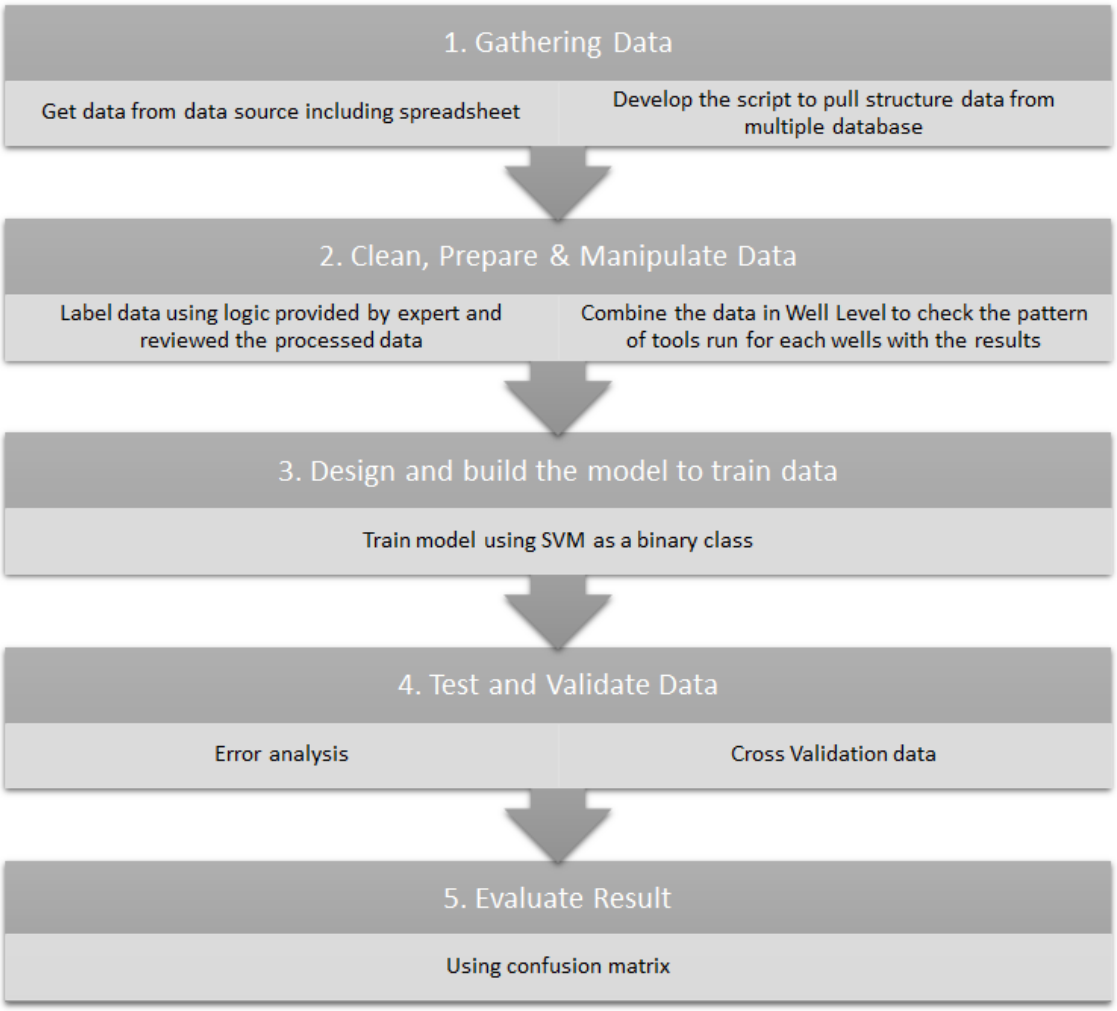

Figure 4.1 Methodology Plan Workflow

4.1 Gathering Data

The historical data that experts accept to use for learning is from 2013-2018 which has approximately 1500 wells run the open hole logging. The criteria for the wells that use for this research need to be the wells in gulf of Thailand which drilled through target depth successfully in production section. There are around 300 wells with failure from historical data.

The data is in different database and in the excel spreadsheet, it needs to be collected prior to preprocessing in the next step. Data from database requires to develop the script to pull data by well level.

4.1.1 Data source

- **WellView** is the database that keep records of drilling data and non-productive events that occurs during drilling
- **OpenWorks** is the database that keep records of geophysical data
- **Wireline Tracking Sheet** is the summarize data by wells of logging tool type used for a well and records of tools running which are not kept directly in specific database
- **Engineering Desktop Tools** has one module to plan and calculate for directional well plan parameter

4.1.2 Data Features

- **Non-Productive Events**
    - Drilling events Type such as lost circulation, well control, cement squeeze from previous section and other well problems.
    - Hours of the non-productive time that occurs before running the open hole wireline logging
- **Drilling Data**
    - The business partner that provide the tools
    - Plan of open hole wireline logging tools type
    - Target depth
    - True vertical depth (TVD)
    - Measure depth (MD)
    - Directional survey from the actual drilling data that were drilled to target depth
    - Drilling fluid rheology
    - Time that use for pulling drill bit and bottom hole assembly out prior to run wireline logging
    - Drilling parameter such as Mud Weight (MW), Rate of penetration (ROP) average at production, Drilling Torque at target depth
- **Geophysical Data**
    - Formation lithology
    - Bottom hole temperature

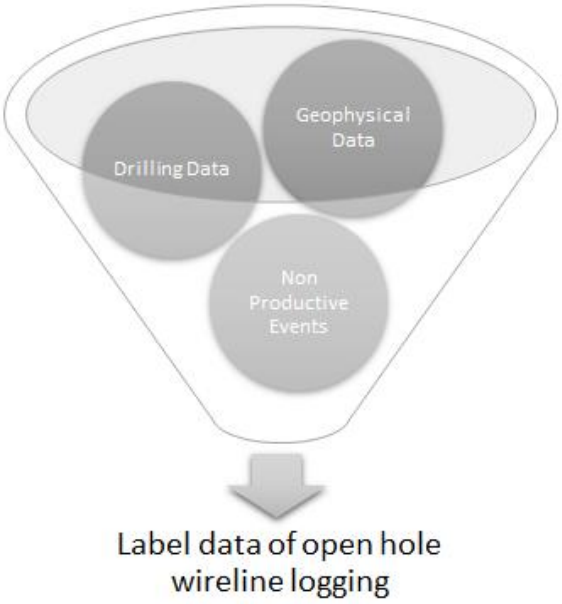

Figure 4.2 Process of gathering source of data and preparing to label the results for classification training

4.2. Clean, Prepare and Manipulate Data

Label data from the historical result of open hole log with the rules of the tool type. Also deal with missing value. In this stage, the record which has missing value will be removed from the observed. The calculation or rule to find the representative value for missing value will be develop later. There are some other steps require to consolidate data into the single format build with rule given from expert.

- **Wireline Tracking Sheet**
  - Formation Tester (RDT, HXPT, HSFT and others), check that the data has Number of Archive > 0 or not. If number of archives is greater than 0 then fail, otherwise is success.
  - Wireline Logging (Quad Combo, Triple Combo), check that the Log to TD information is either Yes means success or No means fail.

However, the data need to be cross check with actual data recorded in the system. The experts help in reviewing the mismatch status of each well.

- **WellView** is a system that has actual activities of logging tool
  - 'LOGWL' or 'FISH' activities with the keyword either 'STUCK', 'SIT DOWN' or 'HUNG UP' mean failed. Using this logic to find out the possibility even they were misspelled or in past tense or passive voice in the comment activities.
  - Dig into the depth that has been record in the comment compare with the plan target depth in the system.

- Check whether it has non-productive time of the unscheduled events

4.3. Design and build model to train data

SVM is selected to train data in the small piece first, data of 2016 to 2018 are tested to build the prediction of two classification results. The scope of this research will use data from 2013 to 2018.

4.4. Validate and testing

The model is validated using the 20% input data which has been prepared earlier to tune hyperparameter. This process aims to find the suitable parameters that bring the model most efficiently and avoid overfitting problem. The model is good with the success result. The failure result will need to be focus and develop algorithm to validate it with more accuracy.

4.5. Evaluate result

The result has been evaluating to find more parameter that has impact to the model to predict 'Fail' class more accurately. The model still using the small set of features, next step is to feed in more features and need to find out the contributing factors that are the major factors to logging result. This is one of the insights that experts would like to know from this project.

5. Objectives

Apply the machine learning technique to get the insight of the data which help making decision prior to running open-hole wireline logging of each well. This could help saving time and cost to avoid the failure of the tools.

6. Scope of Research

6.1 This research focuses on the geological data and well logging information of wells in gulf of Thailand.

6.2 The input of data must be the well that has been drilled to the target depth.

6.3 The output of the model is focus on success or failure of the tools in open-hole condition.

6.4 Failure of the tools in the scope means that the tools is stuck or hung up in the hole either it can fix and pass through the target depth or need to cut the tool and use another tool to pull the failed tool out.

6.5 The failure will not cover the case of tool failure when it works incorrectly or set up with the wrong adjustment and need to update

7. Research plan

7.1 Review

7.1.1 Study the workflow and methodology to drill the wells and the open-hole wireline operation.

7.1.2 Study the concept and condition to identify the failure of the open-hole wireline logging

7.1.3 Study on the related works and research about machine learning technique that use to predict and classify the success or failure

7.2 Prepare data set

7.2.1 Gathering and preparing the data set which include cleaning data, transform data, deal with missing value and convert into the format that can be used to evaluate the proposed method

7.2.2 Consult with the subject matter experts to label the class of prepared data set

7.3 Propose

Select machine technique to apply to the prediction model and design the workflow and experiment prior to build the model

7.4 Implementation

Implement the model following proposed ideas to be used in prediction of two classification results

7.5 Evaluation

Run the implemented model then evaluate and validate the experiment results. Modify some of the parameter to optimize the model.

7.6 Summarize

7.6.1 Analyze the result from the experiment and discuss the outcome with the subject matter expert on drilling operation.

7.6.2 Summarize the results of the research and make a report on this research

Table 7.1 Timeline of this research

Task	Oct-19				Nov-19				Dec-19				Jan-19				Feb-19				Mar-19				Apr-19				May-19			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
7.1.1																																
7.1.2																																
7.1.3																																
7.2.1																																
7.2.2																																
7.3																																
7.4																																
7.5																																
7.6.1																																
7.6.2																																

8. Expected Benefit Gain

8.1 To have a machine learning that learn from historical data and predict the success or failure of the open-hole wireline logging.

8.2 Help improve the decision making

8.3 Help to save cost and time to pull the tools out if the tools are failed.

8.4 Be able to use this framework to apply and find the contributing factors that cause the failure of the open-hole logging.

REFERENCES

VITA

NAME	Maylada Pootisirakorn
DATE OF BIRTH	5 December 1987
PLACE OF BIRTH	Ratchaburi
INSTITUTIONS ATTENDED	Mahidol University Chulalongkorn University
HOME ADDRESS	38/69 Soi Chinnaket Ngamwongwan Road Laksi District Bangkok 10201

1. Cortes, C. and V. Vapnik, *Support-Vector Networks.* Machine Learning, 1995. **20**(3): p. 273-297.

2. Tran, J. *Drilling*. Available from: https://www.studentenergy.org/topics/drilling.

3. Oslo, U.o. *Open hole wireline logging self-learning module*. Available from: www.uio.no.

4. Kleinberg, R., *Well logging overview*. Vol. 13. 2001. 342-343.

5. Tang, Y., et al., *SVMs Modeling for Highly Imbalanced Classification.* IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2009. **39**(1): p. 281-288.

6. Ting, K.M., *Confusion Matrix*, in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G.I. Webb, Editors. 2017, Springer US: Boston, MA. p. 260-260.

7. Wilson, A., *Machine Learning Overcomes Challenges of Selecting Locations for Infill Wells.* Journal of Petroleum Technology, 2018. **70**(10): p. 48-49.

8. Guo, D., et al., *Data Driven Approach to Failure Prediction for Electrical Submersible Pump Systems*, in *SPE Western Regional Meeting*. 2015, Society of Petroleum Engineers: Garden Grove, California, USA. p. 6.

9. Heinze, L. and I.A. Al-Baiyat, *Implementing Artificial Neural Networks and Support Vector Machines in Stuck Pipe Prediction*, in *SPE Kuwait International Petroleum Conference and Exhibition*. 2012, Society of Petroleum Engineers: Kuwait City, Kuwait. p. 13.

10. Serapiao, A.B.S., et al. *Classification of Petroleum Well Drilling Operations Using Support Vector Machine (SVM)*. in *2006 International Conference on Computational Inteligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce (CIMCA'06)*. 2006.

11. Li, Z., et al., *Study on Intelligent Prediction for Risk Level of Lost Circulation While Drilling Based on Machine Learning*, in *52nd U.S. Rock Mechanics/Geomechanics Symposium*. 2018, American Rock Mechanics Association: Seattle, Washington. p. 8.