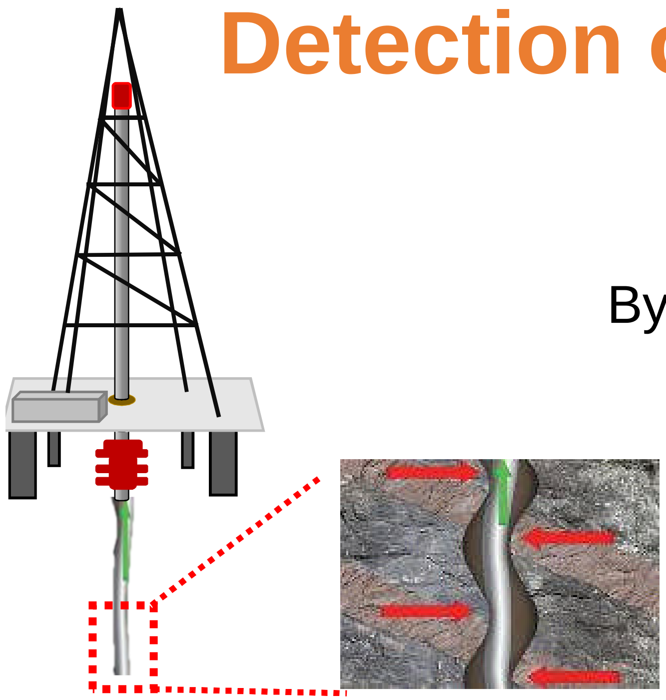
International Conference on Systems and Networks Communications (ICSNC 2020)

Machine Learning and Data Mining for the Early Detection of Stuck Pipe Incidents



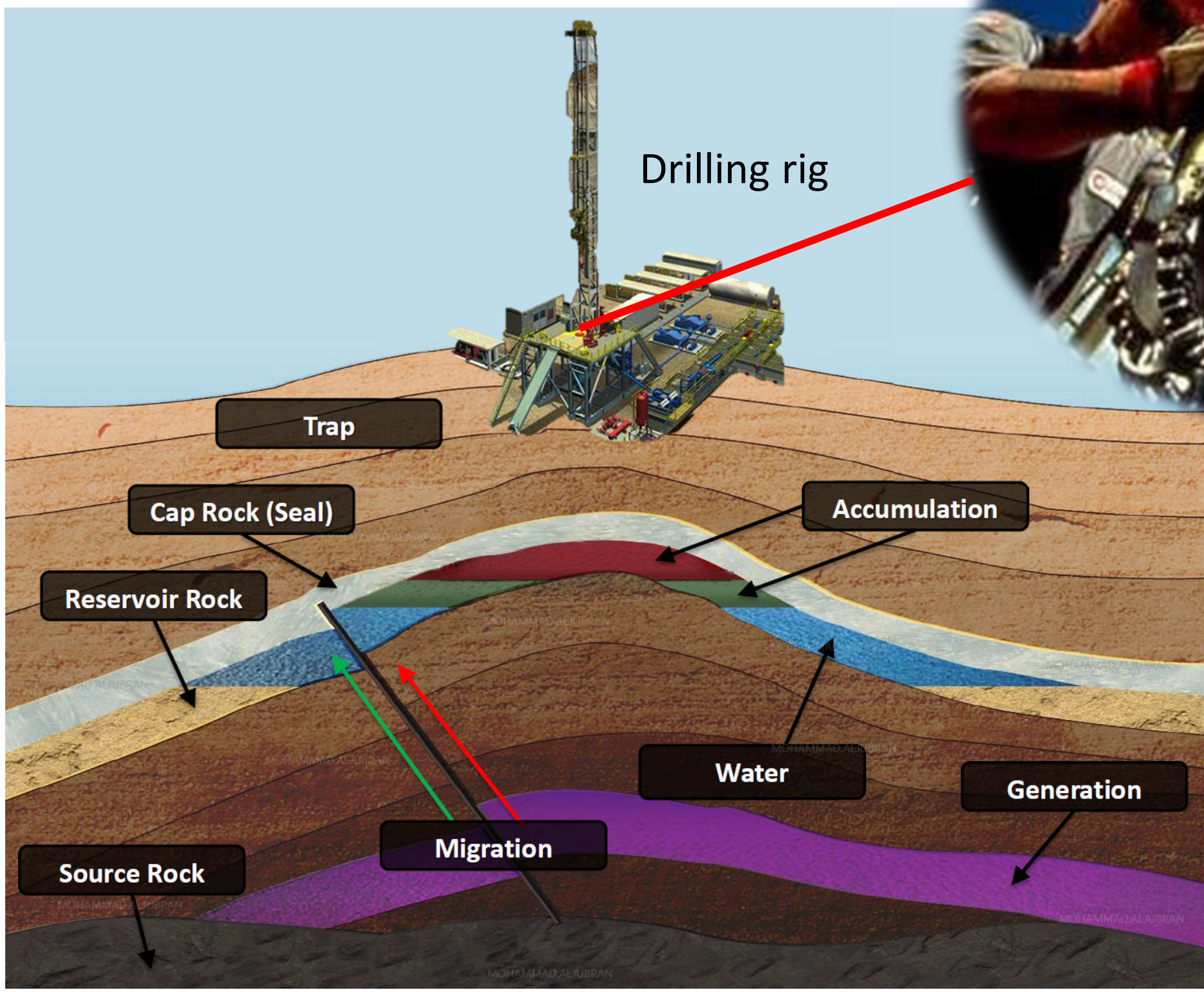
By Arturo Magana-Mora, Mohammad Aljubran

Saudi Aramco

Arturo Magana-Mora

Arturo Magana-Mora obtained his Ph.D. in Computer Science from the King Abdullah University of Science and Technology, Saudi Arabia. During his Ph.D. studies in Computer Science and a postdoctoral fellowship at the National Institute of Technology (AIST) in Japan, he developed novel artificial intelligence models to address problems in biology, genomics, and chemistry that resulted in several peer-reviewed publications in high-impact journals, poster presentations, and invited talks. Currently, he works at the EXPEC Advanced Research Center, Saudi Aramco, where he has opened up many new opportunities in the domain of Drilling Automation and Optimization and catalyzed existing work. During his career he has used his expertise in computer science to bridge artificial intelligence with biology, genomics, chemistry, and the oil and gas industry, and currently serves as Guest Editor and referee for several scientific journals.

Drilling Background



Objectives

- Reach hydrocarbons/gas reservoirs
- Different depths

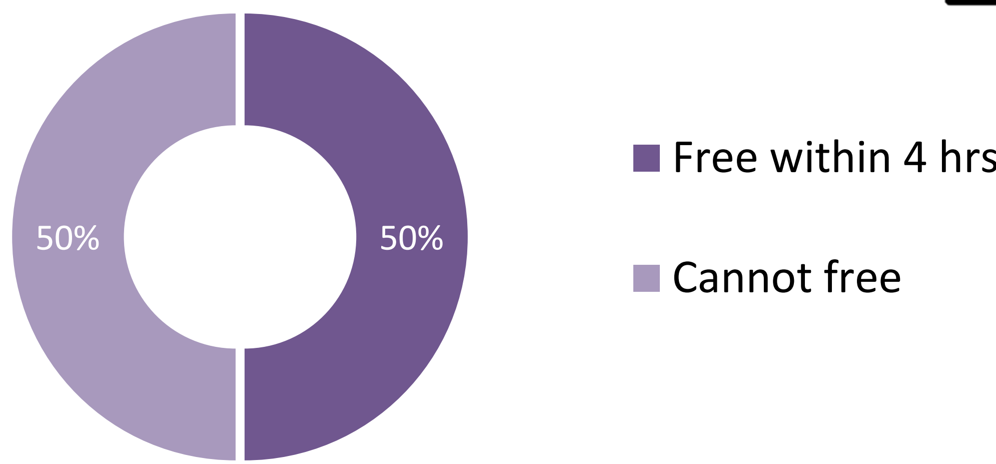
Considerations

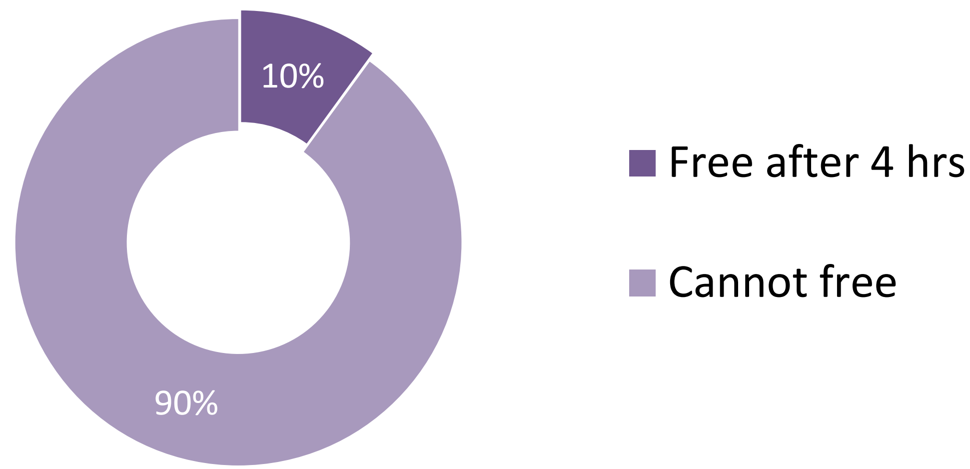
- Hydrostatic pressure
- Hole cleaning
- Well control
- …

Business Impact

Non productive time

- Stuck pipe
- Rig or equipment failure
- Downhole tool failure
- Hole problems (enlargement/collapse)
- Slow rate of penetration
- Among others

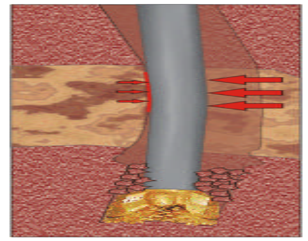




Background

Differential Sticking

- Permeable Formation
- Filter cake
- Overbalance
- Lack of pipe movement

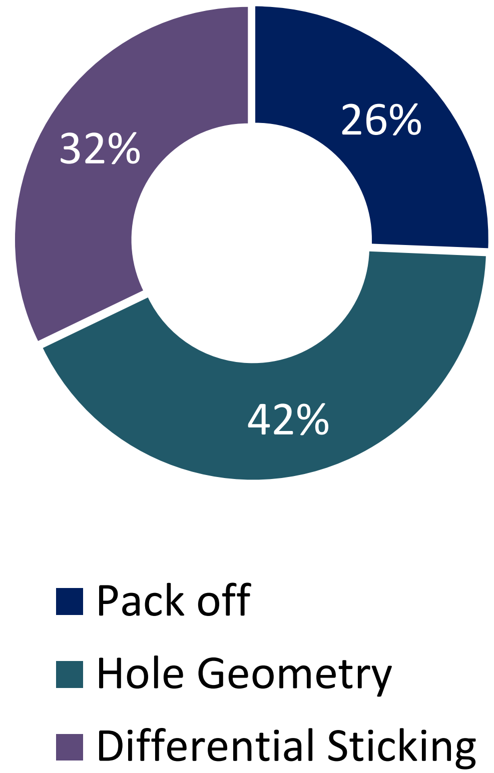


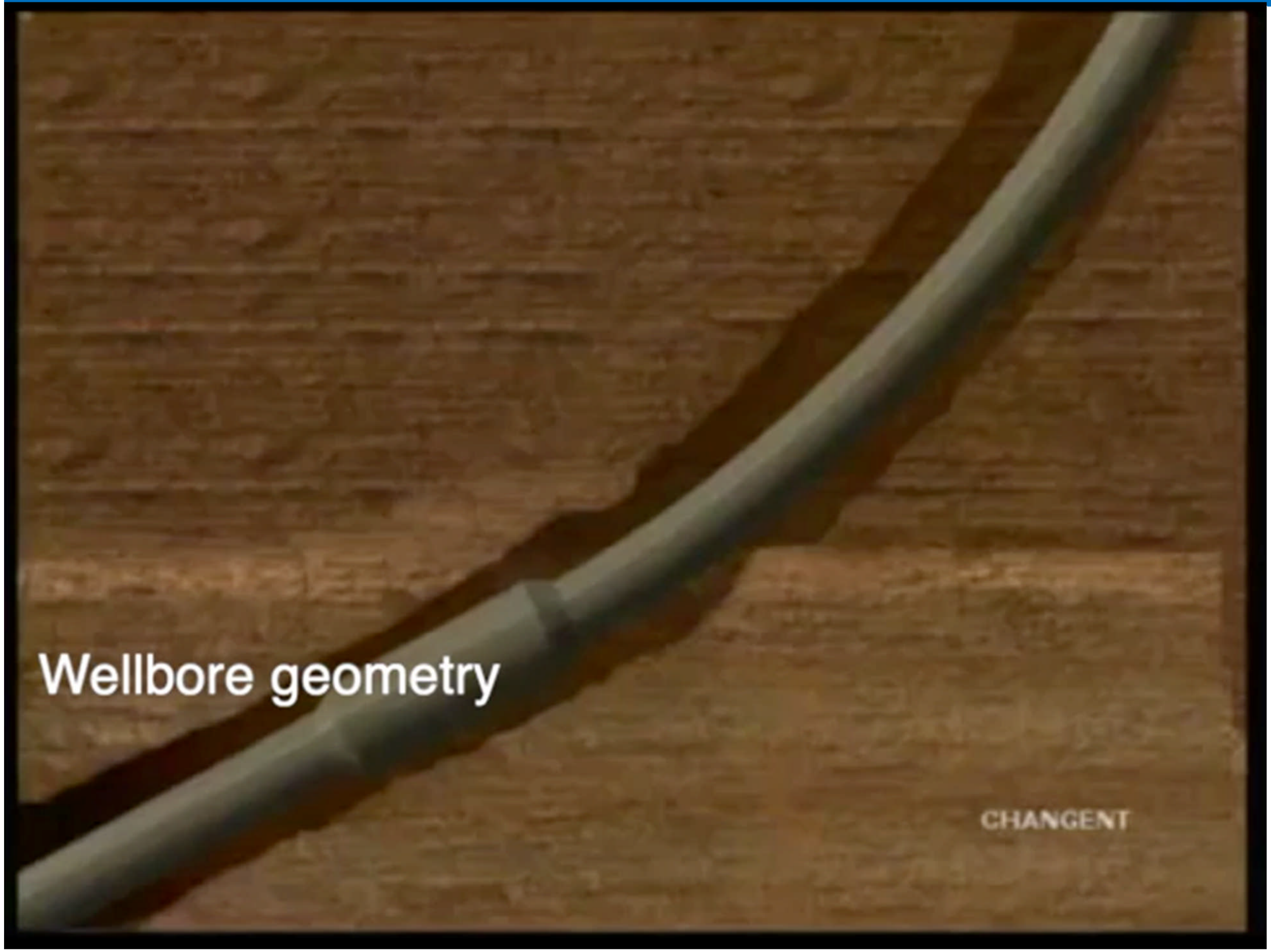
Pack-off/Hole Instability

- Hole cleaning
- Formation related

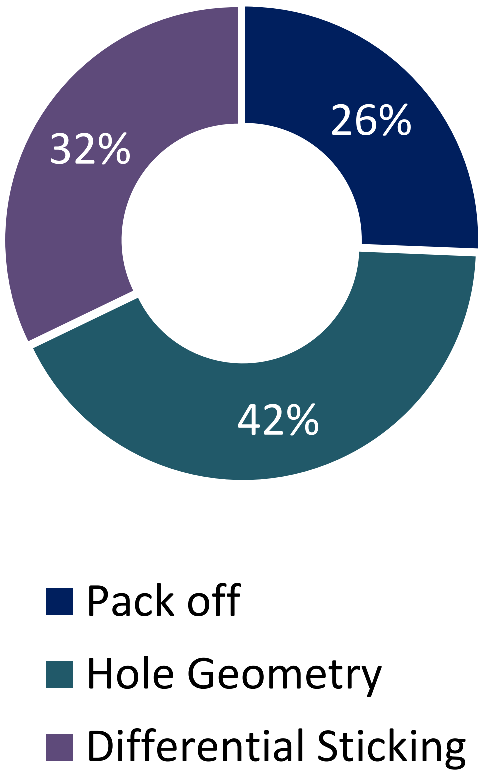
Wellbore Geometry

- Doglegs and Micro doglegs
- Ledges
- Under-gauged holes
- Mobile formation





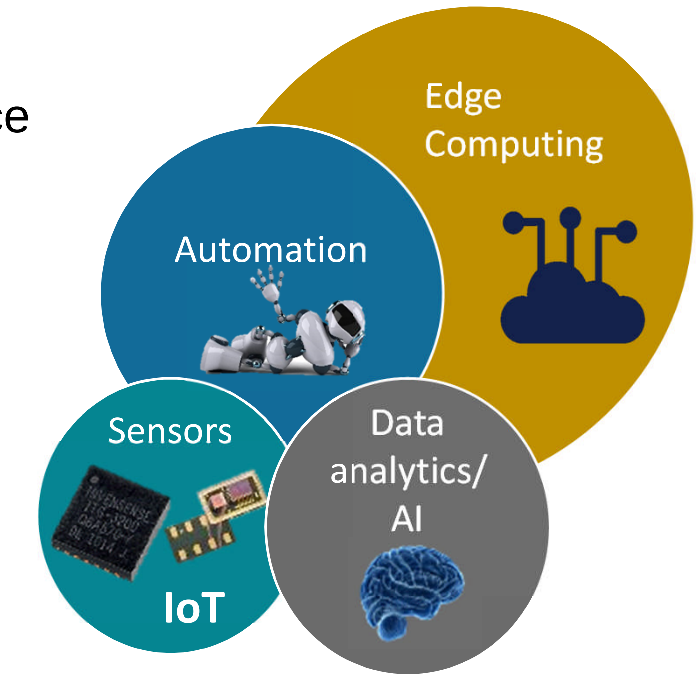
Source: Youtube - Changent



Background

What can we do to enhance this framework.

Optimal performance



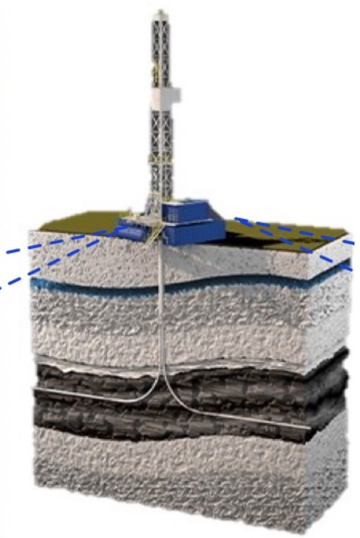
Background

- Discriminant Analysis Technique (Hampkins et al.,1987).
- Multivariate Discriminant Analysis (Biegler and Kuhn, 1994).
- Artificial Neural Network (Siruvuri, 2006; Miri et al., 2007).
- Adaptive Neuro Fuzzy Logic (Murillo et al., 2009; Naraghi et al., 2013).
- Support Vector Regression (Jahanbakhshi et al., 2012).
- Case-based Reasoning (Sadlier et al., 2013; Ferreira et al. 2015).
- Support vector machines (Alshaikh et al., 2019).

Methodology – Data acquisition

Sensors measure different operational parameters, e.g., revolutions per minute (RPM), torque (TQ), hook load (HL), etc.

$$t_1 = \{RPM_1, TQ_1, HL_1 \ldots\}$$
$$t_2 = \{RPM_1, TQ_1, HL_1 \ldots\}$$
$$t_3 = \{RPM_1, TQ_1, HL_1 \ldots\}$$
$$t_4 = \{RPM_1, TQ_1, HL_1 \ldots\}$$
$$t_5 = \{RPM_1, TQ_1, HL_1 \ldots\}$$
$$t_6 = \{RPM_1, TQ_1, HL_1 \ldots\}$$
$$\ldots$$
$$t_n = \{RPM_1, TQ_1, HL_1 \ldots\}$$



Frequency = 1 – 0.2 Hz

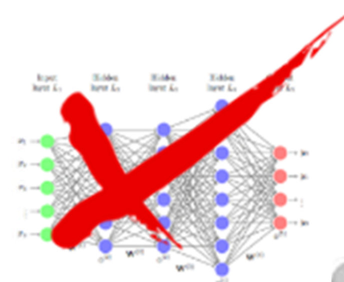
Using few wells for training may not be optimal…

Big amounts of data (time steps) but not necessarily meaningful.

Why? Learning data from **few** wells may not produce **generalizable** models.

Weather forecast

Years
2000 & 2005

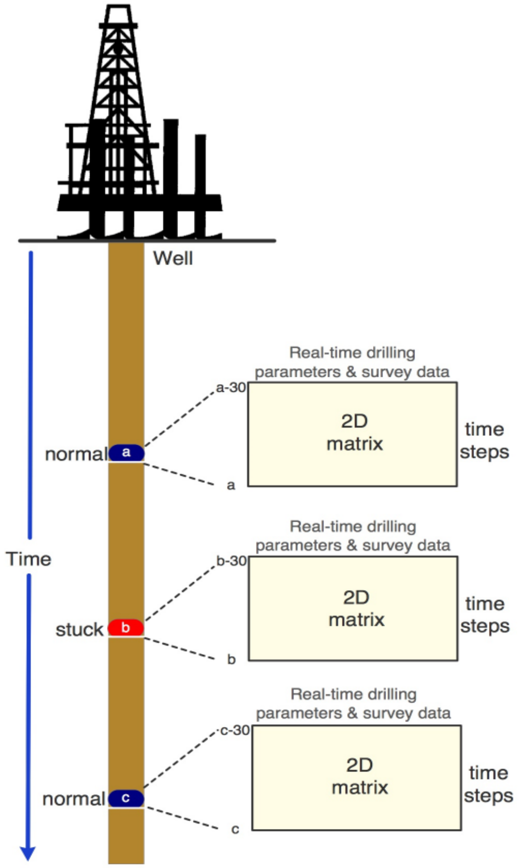


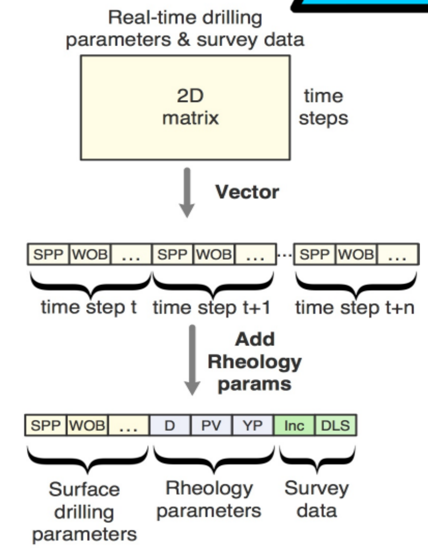
Methodology - Data acquisition

- Historical stuck pipe incidents during drilling.
- 30 minute window before the stuck pipe incident (663 parameters).

Type	Parameter	Units
Real-time surface drilling parameters	Hook height (HKHT)	ft
	Weight on bit (WOB)	klbf
	hook load (HKL)	klbf
	Stand pipe pressure (SPP)	psi
	Torque	kft.lbf
	Flow-in rate	gpm
	Flow-out rate	0-100 (%)
	Rate of penetration (ROP)	ft/h
	Revolutions per minute (RPM)	rpm
Rheology parameters	Density	PCF
	Yield point (YP)	numeric
	Plastic viscosity (PV)	numeric
Real-time survey data	Inclination	degrees
	Dogleg severity	degrees/100 ft

Note: Surface drilling parameters are referenced to time (0.2 hertz)





Statistical number of wells with stuck incidents.

1:3 stuck/normal ops. Ratio.

Methodology - Data Normalization

- Normalized using the z-score in order to compensate for the differences in the ranges of data for each parameter per case.

$$z_i = \frac{x_i - \mu_i}{\sigma_i}$$

- Computed the analysis of variance (ANOVA) F-value to rank the relevance of the features.

- Identified the key parameters to be used in further model development.

Methodology – Machine Learning Models

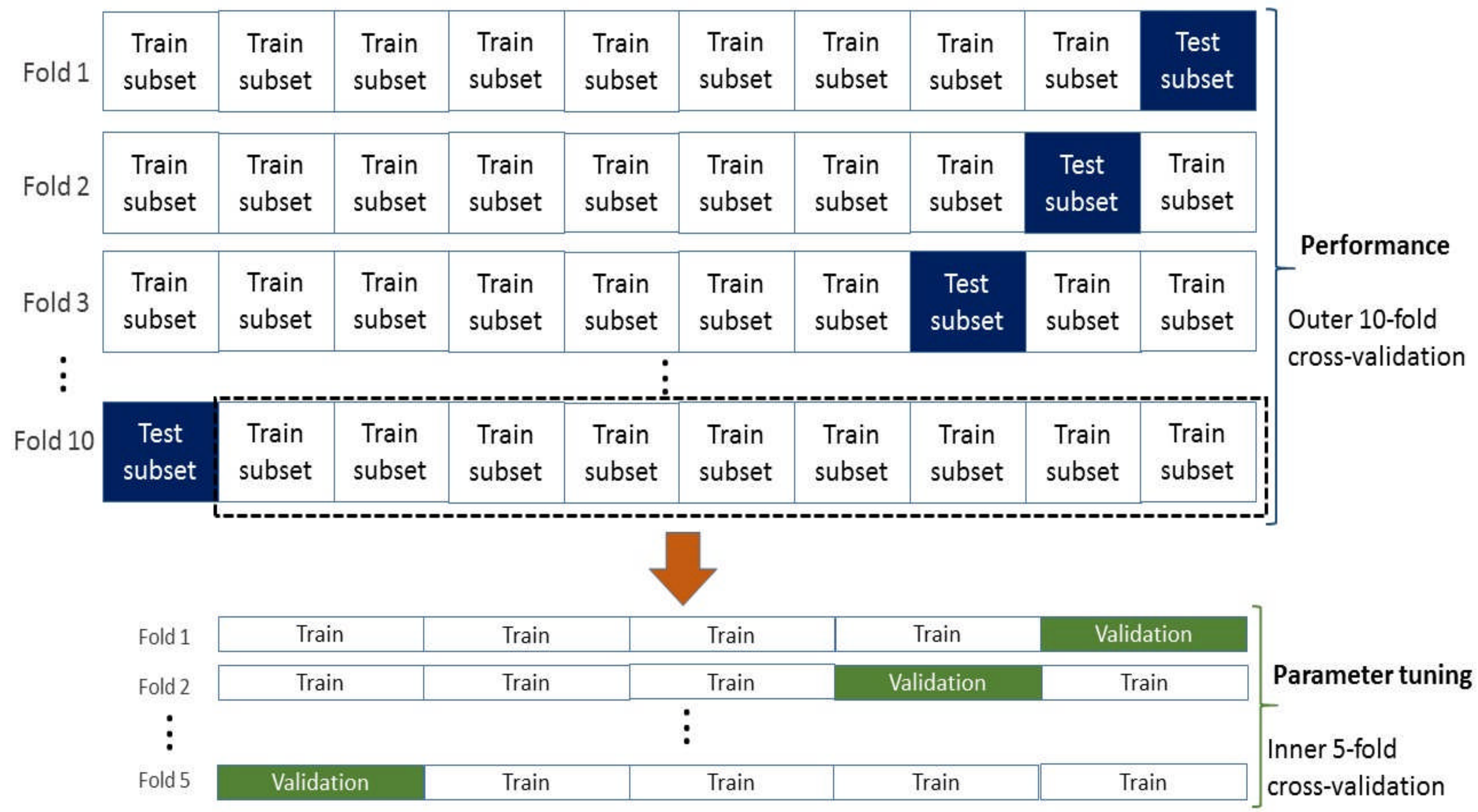
- Four robust machine-learning algorithms were used to derive models for the detection of stuck pipe incidents:
    - **Random Forest (FT).**
    - **Artificial Neural Networks (ANNs).**
    - **Decision Trees (DTs).**
    - **Support Vector Machines (SVMs).**

- The results of each method were then compared for accuracy in detection.

Methodology – Model Training & Testing

Nested cross-validation

- Used 10-fold cross-validation to test models performance.
- Used 5-fold cross-validation for tuning parameters.



Methodology – Performance Measures

Measure	Equation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$

Avoid false positives.

Focus on the **precision** of the model.

Results & Discussion

Feature ranking based on the analysis of variance (ANOVA) F-value

Average Rank	Parameter
1	Rate of penetration (RPM)
2	Stand pipe pressure (SPP)
3	Flow-in rate
4	Weight on bit (WOB)
5	Hook height (HKHT)
6	Hook load (HKL)
7	Torque
8	Rate of penetration (ROP)
9	Dogleg severity
10	Inclination
11	Flow-out rate
12	Mud yield point (YP)
13	Mud plastic viscosity (PV)
14	Density

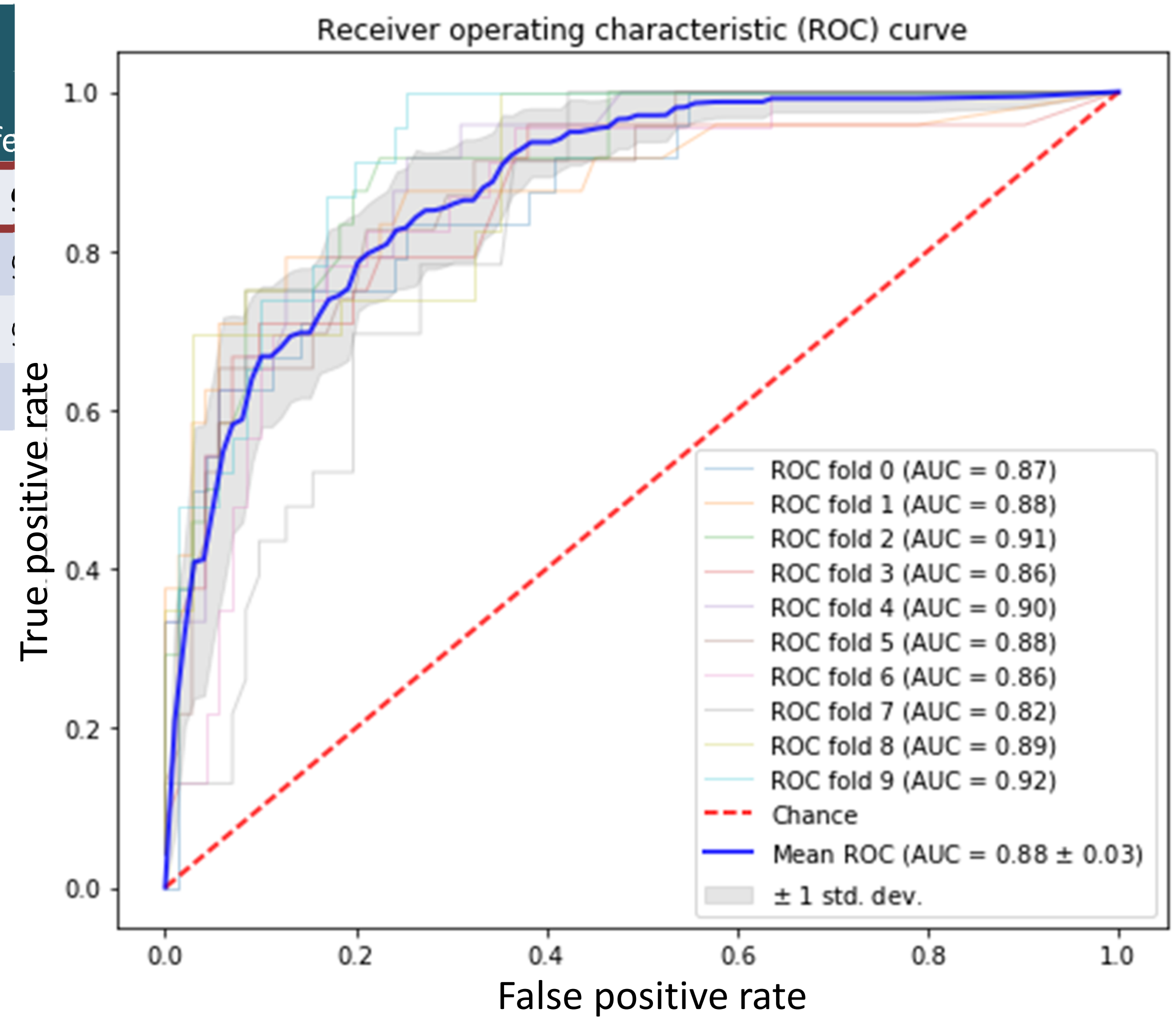
The table shows the considered parameters based on the averaged F-value for all time steps to avoid listing the 663 parameters.

Results & Discussion

Model	Accuracy (%)		Precision (%)		Sensitivity (%)		
	All features	Subset 66%	All features	Subset 66%	All features	Subset 66%	fe
RF	**81.80**	**82.75**	**79.32**	**75.94**	**60.00**	**55.43**	**9**
DT	80.10	80.10	64.11	62.70	47.66	50.21	9
ANN	80.00	80.52	64.17	65.66	44.27	45.92	9
SVM	75.13	75.13	0	0	0	0	

Radom forest:
80 precision (~20 % false alarms).
60 sensitivity (60% stuck incidents detected).

Note that class-weighted SVM is designed to deal with unbalanced data by assigning higher misclassification penalties to training instances of the minority class. However, the model was unable to learn the dependencies of the features in the vector of 663 features.



Averaged ROC from an RF using 10-fold cross-validation.

Conclusions & Steps Forward

- **Large-scal**e analysis of a statistically representative number of wells with stuck pipe incidents.
- Real-time surface **drilling** parameters, **survey** data, and **rheology** parameters for **30 minutes** before the stuck pipe incident took place (time series analysis).
- Performed **analysis of variance F-value** to reduce the feature set from 663 to 438 (feature selection).
- RF achieved the best performance with a **precision score of ~80%** and **sensitivity of ~60%** from a 10-fold cross-validation.

Current efforts and ongoing work

- Data cleansing and labeling.
- Feature engineering (wavelet transformations, rig activities, survey data).
- Deep learning models, i.e., long short term memory (LSTM) neural networks, convolutional neural networks, among others.

Thank You!