

IoT-Driven Big Data in Construction Drilling Tools

—

Base Material Detection and Wear Estimation

Stefan Goldener

Widen 3

CH-9473 Gams

150232

Master's thesis

to obtain the degree of

Master of Science in Information Systems

Confidential

University of Liechtenstein

Graduate School

Programme: Information Systems

Module: Data Science

Course: Master's thesis

Assessor: Dr. Johannes Schneider

Co-Assessor: Prof. Dr. Stefan Seidel

Working period: 01/03/2016 to 30/06/2017

Date of submission: 30/06/2017

Table of contents

Confidentiality clause (Sperrvermerk)	iii
Abstract	iii
1 Introduction and Motivation	9
2 Research Gap	9
3 Literature Review	10
3.1 Tool	11
3.2 Core and Drill bit.....	11
3.3 Tool - Drilling and Cutting.....	11
4 Research Question	12
5 Method	13
6 Theories and Concepts	15
6.1 Supervised Learning.....	15
6.2 Semi-Supervised /Unsupervised Learning	15
6.3 Definition of Big Data	16
6.4 Definition of IoT	17
6.5 Data Mining Methodologies and Process Models.....	18
6.6 Processing Big Data	19
7 Introduction to Content of the Main Part of the Thesis	19
8 Data Collection, Retrieval, and Transformation	20
8.1 Data Collection.....	20
8.2 Data Generation.....	20
8.3 Data Content.....	21
8.4 Data Retrieval.....	22
8.4.1 Extract	22
8.4.2 Transform.....	23
8.4.2.1 Binary Data.....	23
8.4.2.2 Aggregation/Filtering	23
8.4.3 Storing Data.....	24
9 Data Exploration	25
10 Data Preparation	26
10.1 Data Filtering, Smoothing & Interpolation	26
10.1.1 Filtering Algorithms	26

10.1.2	Smoothing Algorithms	26
10.1.3	Approximation and Interpolation Algorithms	26
10.1.4	Comparison of Smoothing, Filtering & Interpolation Algorithms	26
10.1.4.1	Hampel – Filtering Algorithm	26
10.1.4.2	Approx – Interpolation Algorithm.....	27
10.1.4.3	Spline - Interpolation Algorithm	27
10.1.4.4	Smooth – Smoothing Algorithm	28
10.1.4.5	LOESS and LOWESS - Smoothing Algorithm.....	28
10.1.4.6	Smooth Spline – Interpolation & Smoothing Algorithm.....	28
10.1.4.7	Visual Comparison Overview – Single Hole	29
10.1.4.8	Visual Comparison Overview – Multiple Holes	29
10.1.4.9	Performance – Multiple Holes.....	30
10.2	Identified Challenges with Filtering.....	31
10.3	Data Correction	32
11	Feature Engineering.....	33
11.1	Forward Speed.....	33
11.1.1	Derivative Approach	33
11.1.2	Segmented Linear Interpolation	33
11.1.3	Segmented/Piecewise Regression Approach.....	36
11.1.4	Discussion Forward Speed	38
11.2	Feed Force Difference	38
11.3	Pressure & Mechanical Stress	38
11.4	Density	38
11.5	Friction	39
11.6	Acceleration as a Feature	40
11.6.1	Transformation to the Frequency Spectrum.....	40
11.6.2	Standard Deviation as Feature.....	40
11.7	GPS Data as a Feature	41
11.8	Cell ID as a Feature	42
11.9	Gear Information as a Feature	43
11.10	Energy in Rotary Drilling with Hollow Drill	44
11.11	Young's Modulus.....	47
11.12	Penetration per Revolution.....	47

12	Unsupervised Labeling with an Unsupervised Algorithm.....	48
12.1	Variable Analysis	48
12.1.1	Variable Plot.....	48
12.1.2	Principle Component Analysis	49
12.2	Choosing the Number of Clusters	50
12.3	K-Mean on Subset Data	51
12.4	Unsupervised Clustering on External Data	53
12.4.1	K-Mean.....	53
12.4.2	K-Median.....	53
12.4.3	K-Medoids.....	53
12.4.4	Expectation–Maximization (EM) Algorithm	53
12.4.5	CLARA (Clustering Large Applications)	53
12.4.6	Unsupervised Random Forest.....	54
12.5	Performance Comparison – Unsupervised Algorithms.....	54
12.6	Problems with Unsupervised Learning	55
12.7	Second Unsupervised Approach after the Supervised Approach.....	55
13	Hand Coding of Data	56
13.1	Creation of an Interface.....	56
13.2	Assisted Labelling	57
13.3	Data Split.....	59
14	Supervised Models.....	61
14.1	Neural Network	61
14.2	Deep Neural Network.....	62
14.3	Decision Tree	63
14.4	Boosted Decision Tree	63
14.5	Extreme Gradient Boosted Tree (xgbtree)	64
14.6	Random Forest	64
14.7	Supervised Models - Performance.....	65
14.7.1	BMC - Model 1R and Model 1 Comparison.....	66
14.7.2	Special Training Set - Test for Overfitting in the First Model	67
14.7.3	BMC Model 2 and Model 3 Comparison	68
14.7.1	BMC Model 4 – GPS Cluster Importance Test.....	70
14.7.2	BMC Model 5 – Implementation of Findings of Model 2&3	71

14.7.3	BMC Model 6 – Core Bit Estimation Test and new Features	72
14.7.4	BMC Model 7 – Implementation of Findings of Model 6	73
14.7.5	BMC Model 8 – Enrich Missing Data.....	74
14.7.6	BMC Model 9 – Implementation of all Findings	75
14.7.7	Supervised Algorithms Summary - Comparison.....	76
14.8	Discussion on Supervised Learning	78
14.9	Minimum Amount of Drilled Minutes	79
14.10	Counting the Number of Reinforcement Bars.....	80
15	Deployment of the Base Material Classification Model.....	80
15.1	Real-Time Classification.....	80
15.2	Near Time Feedback	81
16	Wear Estimation.....	83
16.1	Additional Internal Data Source for Wear Estimation	84
16.2	Material Classification for Wear Estimation Model.....	85
16.3	Regression Performance Measure	86
16.3.1	NRMSE	86
16.3.2	Mean absolute error (MAE)	86
16.4	Wear Model Development	87
16.4.1	Wear Model 1 – All Data Points	87
16.4.2	Wear Model 2 – Minimum of 0.04mm Wear.....	89
16.4.3	Wear Model 3 – Minimum of 0.1mm Wear.....	91
16.4.4	Wear Model 4 – Variable Selection	92
16.4.5	Lasso Algorithm for Wear Estimation	94
16.4.6	Ridge Model for Wear Estimation	96
16.4.7	Elastic-Net for Wear Estimation	97
16.5	Discussion - Wear Estimation	98
17	Discussion.....	99
18	Conclusion.....	100
19	References	101
	List of figures	111
	List of tables.....	113
	List of abbreviations.....	115
	Appendix A	117

Appendix B	119
Declaration of authorship	130