Sandman Requirements Definition Document

**Project Name: Sand Parameters Prediction**

**Document details:**

|  |  |
| --- | --- |
| **Tag** | **Details** |
| Feature version | Version 1.0 |
| Document version | Version 1.0 |
| Document Author | Kaaviya |
| Last edited on | 2nd April, 2020 |
| Last edited by | Kaaviya |

# 

# Purpose of the document

The purpose of this document is to provide **teams** working on Sandman with the necessary and required information for them to commence work on the task at hand.

# Document history

|  |  |  |  |
| --- | --- | --- | --- |
| **Update made on** | **Update made by** | **Cause** | **Affected section** |
| **01-10-2019 to 02-01-2020** | **Kaaviya** | This task was already done by an intern and now it will be done by me with some in-depth analysis. |  |
| **25-02-2020** | **Kaaviya** | Brakes data is taken and pre-processing work is completed i.e the work such as treating null values, treating outliers etc.. |  |
| **26-02-2020** | **Kaaviya** | Prediction by having GCS as output is completed. 23 regression algorithms are applied with base hyper parameters. R-Squared comes to 60% approximately. All the best models will be taken and tuned to get better results . |  |
| **26-02-2020** | **Kaaviya** | Analysis by having compactability as output is completed. The best model produced an R-squared of 75% with 0.71 as RMSE |  |
| **26-02-2020** | **Kaaviya** | Analysis by having Active Clay as output is completed. The best model produced an R-squared of 96% with 0.010 as RMSE. |  |
| **02-03-2020** | **Kaaviya** | Analysis by having WTS as output is completed. |  |
| **13-03-2020** | **Kaaviya** | The initial coding part is completed but there will be many iterations of models which will be ran again and again until best result is achieved. |  |
| **23-03-2020** | **Kaaviya** | Some of my Sand parameters had a very high RMSE value and R-squared values. The main objective was to reduce the RMSE and increase the R-squared value. I started with GCS . Before doing hyper parameter tuning I tried to figure out which sand parameter is increasing my error and decreasing my accuracy. So I tried removing each sand parameter and kept running the model again and again . Later I did the same from backwards. I tried adding the sand parameters one by one to see if any parameters has any positive impact on my results. I did not succeed in this attempt. By doing this I had a very small impact in my error. Eg: if my error was 18.9% it came down to 18.7% and if my R-squared was 60.3 it increased to 60.5. Then with the help of feature importance I found only those features which contributes to my model and I tried running a model with only those features. Still my error was High. I was doing the same with other sand parameters. I do find some positive trends but that still has a huge error.There are still few more sand parameters to deal with and few more methods that is left to try. |  |
| **24-03-2020** | **Kaaviya** | A analysis was done and a report was prepared on the analysis and findings of GCS and WTS and the methods that did not work . |  |
| **25-03-2020** | **Kaaviya** | A analysis was done and a report was prepared on the analysis and findings of compactability and LOI and the methods that did not work . |  |
| **26-03-2020** | **Kaaviya** | Analysis was done on rest of the sand parameters for brakes and methods that did not work were noted. |  |
| **27-03-2020** | **Kaaviya** | The error rate was high  for some of the models. The accuracy is also less. Tried different methodologies to improve the result but after a certain extent the model tends to overfit. |  |
| **30-03-2020** | **Kaaviya** | For the foundry brakes, adjusted R-squared is calculated and it is compared with R-squared along with the bias and variance error to see if the model tends to overfit or underfit. |  |
| **30-03-2020** | **Kaaviya** | Munjal Base model building is completed for 4 sand parameters. |  |
| **31-03-2020** | **Kaaviya** | As of now Model building is completed for Munjal and Maxwell. Only the best models which was winning in brakes were ran and the results are recorded. These models will be ran again again in various iterations until best result is achieved. |  |
| **01-04-2020** | **Kaaviya** | There are certain parameters were different model seems to work well. Fine tuning, checking VIF, performing RFE works on those models were done to arrive at the best results., and the sand parameter report has been updated with the results. |  |

# 

# New Feature Request / Enhancement Request

# Preliminary details \*

### Please provide a high-level description of the requirement as either a problem that is faced by the foundry in the context of day-to-day foundry operations and/or a potential pain-point with regards to the usage of the application.

**It is not possible for the foundries to measure all the sand parameters on daily basis. If a regression model is prepared for all the sand parameters with good score then the analytics team will be able to predict those values even if it is not measured.**

### Please provide details of how this problem is currently or has traditionally been handled by foundry personnel.

**Currently if a particular sand parameter is not measured then those sand parameter columns will be left blank without any data.**

### Does the proposed work pertain to development of a new feature? (Answer Yes/No)

### If you answered No to the above question, does it then pertain to the enhancement of an existing feature? (Answer Yes/No)

**Yes, If the model that is going to be built is working as per the requirement then this will be implemented which in turn leads to the development of a new feature.**

### Was this feature directly requested by the end user? (Answer Yes/No)

### If you answered Yes to the above question, please provide the requirements (verbatim) as communicated by the end user

### If you answered No to the above question, please provide the rationale / train of thought through which this requirement was envisioned by MPM personnel

**No, it was not requested by the end user.** **Null values has been a major problem in most of the analysis. There were Non-Availability of data in many of the rows and the foundries are not able to provide with the data since it was not measured. If a regression model which** **predicts sand parameters effectively then predictions with those data can be made effective** **with accurate results.**

In which of the following areas of foundry operations does the proposed feature/enhancement create value addition?

|  |  |
| --- | --- |
| **Foundry Operations** | **Value Addition (Yes/No)** |
| **Data** | |
| Data Acquisition | **Yes** |
| More number of data points | **Yes** |
| Data Validation | **Yes** |
| **Decision Support** | |
| Prediction of raw material consumption | **No** |
| Prediction of product quality | **No** |
| Prediction of equipment failure | **No** |
| **Optimization** | |
| Optimization of raw material consumption | **No** |
| Optimization of product quality | **No** |
| **Planning/Scheduling/Resource Management** | |
| Optimization of production plan | **No** |

### For each of the identified areas of foundry operations where value addition is expected, illustrate with at least 1 scenario for each area, the manner in which this value addition is achieved.

### 

**If more data points are there then the regression model will be exposed to lot of scenarios and therefore the results will be more accurate. If any column has more than 50% or 80% of null values then that entire column will be dropped from the analysis. Sometimes important columns will also be dropped which in turn affects the performance of the model. The data is taken from the client to perform analysis so data acquisition adds value.**

# 

# Project Timeline\*

### Describe in detail in a step-by-step manner, how the defined problem statement will be solved

* The Entire Analysis will be based on any 3 foundries. I have chosen those foundries to be Brakes, Munjal and Maxwell.
* First Brakes analysis will be done.
* The prep-processing work includes grouping the data based on ‘Date, Component ID, and Shift, treatment of null values and treatment of outliers.
* Then a base model will be built with 24 regression algorithms to see which model performs the best.
* Then the best model is taken and tuned with the help of GridSearchCV.
* If GridSearchCV doesn’t work, ensembled methods such as Bagging and Boosting will be applied to the model to see if there are any improvement in the results.
* A multi collinearity test will be done by analysing the sand parameters with the help of variable inflation factor. Any sand parameter with a VIF value of more than 5 will be removed and the regression models will be ran again to see if there are any improvement in the results.
* Recursive feature elimination method will be applied on the best model. This RFE methods will rank the attributes based on how much it contributes to the model. Based on the RFE results the sand parameters which are ranked high will be kept as an input and all the best regression model will be built again to see if there are improvement in the results.
* Any regression model which provides the best results by applying the above methods will be considered as the best model.
* Training and Test data for R-squared and Adjusted R-squared will be recorded to see if the model overfits or underfits.
* After recording the best results for brakes then analysis will be done for Munjal Kiriu.
* Pre-processing work will be same as how it was done for brakes. Initially only the winning regression models in brakes will be built in munjal to see whether the same regression models are winning in Munjal as well. If that is not the case then all the 24 regression models will be built again to figure out the best model.
* Similarly, the above same process will be done for Maxwell. Initially only the winning regression models in brakes will be built to see if the same models are winning in Maxwell as well. If that is not the case then all the 24 regression models will be built again to figure out the best models.
* All the results will be recorded along with train and test data for r-squared and adjusted R-squared.
* Finally, the results of all the 3 foundries will be analyses. There is no problem if a model is consistently winning across all the 3 foundries with good R-squared and less error. If the winning model is different across all the foundries then it should be looked at.
* If any sand parameter is less across all the foundries then that should be looked and analysed as to what can be made differently.
* If a sand parameter provides good results in one foundry but does not perform well in another foundry then that should also be looked at.

This entire project is expected to be completed in the first week of April.

# Available Data \*

### Data files

Please enter the details of the data files one below another for each file that will be provided

#### File meta-data

|  |  |
| --- | --- |
| **Tag** | **Value** |
| Name of the file | Prepared Sand Data – Brakes, Munjal, Maxwell |
| High level description of what the file contains | It contains all the sand parameters such as gcs, active clay, wts, loi, moisture, permeability etc.. |
| Last modified on | April 2nd ,2020 |
| Last modified by | Kaaviya |

#### 

#### Field details sec. A

|  |  |  |
| --- | --- | --- |
| **Name of the column** | **Description** | **Location in the schematic where it is measured** |
| **GCS** | **Sand Parameter** | **Foundry** |
| **Compactability** | **Sand Parameter** | **Foundry** |
| **Active Clay** | **Sand Parameter** | **Foundry** |
| **WTS** | **Sand Parameter** | **Foundry** |
| **LOI** | **Sand Parameter** | **Foundry** |
| **Moisture** | **Sand Parameter** | **Foundry** |
| **Total Clay** | **Sand Parameter** | **Foundry** |
| **Inert Fines** | **Sand Parameter** | **Foundry** |
| **Volatile Matter** | **Sand Parameter** | **Foundry** |
| **Permeability** | **Sand Parameter** | **Foundry** |
| **Shear Strength** | **Sand Parameter** | **Foundry** |
| **Temp of Sand Before Mix** | **Sand Parameter** | **Foundry** |
| **Temp of Sand After Mix** | **Sand Parameter** | **Foundry** |
| **Ambient Temperature** | **Sand Parameter** | **Foundry** |
| **GFN/AFS** | **Sand Parameter** | **Foundry** |
| **Relative Humidity** | **Sand Parameter** | **Foundry** |

### **Additional information**

Please enter additional information that you wish to provide like **ASSUMPTIONS / ANY CONSTRIANTS etc..**

# Acceptance Criteria \*

### Please state the average acceptable deviation of the algorithm’s output from the expected output as a percentage.

**An algorithm with an output greater than 75% is considered to be a good one. Anything less than 75% will be analysed carefully to understand the behaviour of the model and different methodologies should be tried in order to achieve 75%.**

### Based on your response to earlier questions, please have domain personnel provide some inputs and expected outputs (either as a number or as a trend). These will be used as initial test-cases which shall be used for validation of the algorithm.

### 

### 

# Appendix

## **Note on Mandatory sections**

Sections which are **highlighted,** and which contain an asterisk **\*** indicate that the information that is to be provided therein is mandatory. The agreed upon protocol is that, without providing said information **teams** are not expected to commence work on the task.

## **Editing guidelines**

It is understood that during the course of refining the requirements, this document would undergo changes. Some guidelines have been provided to assist personnel in this regard.

1. Every change that is made to the original requirement is expected to be captured in this document.
2. Such updates are expected to be made prominent. Good practice would be to
   1. Highlight the section that has been updated with new information like so
   2. Append the update to the appropriate section of this document with the keyword **“Update”** in the following pattern

**<Update><person who performed the update><date-time when update was performed>**

1. Insert an entry into the document history table at the start of the document. This table is expected to contain history of changes made in chronological order.
2. Only personnel who raise the requirement request should perform the edits. The **teams** are neither expected nor allowed to perform these edits
3. Follow the template provided by the document. After performing edits try to view the document outline to ensure consistency