# Motec Data input information Study

## Without traction, torque and speed

To try to see what information was useful to include with predicting the density of tailing using the data collected by running the machines I used a linear regression machine learning model to try and find useful data.

To start with i assumed traction, torque and speed would be useful so I removed them from input data along with the lat, long, data and time information. I used the information provided by David yang to Categorized the data into 5 groups.

Here are the coefficient and accuracy data from the linear regression model

Number of rows in the training data: 217

Linear Regression Model Performance:

* Mean Squared Error: 0.351929297474935
* R-squared: 0.8048620137728084

Feature Weights (Coefficients):

* Sum of Air\_Temp\_Manifold 4.441009e-02
* Sum of Ambient\_Air\_Temp -4.724941e-02
* Sum of BEARING\_PRESSURE\_LEFT\_IDLE -3.586642e-01
* Sum of BEARING\_PRESSURE\_RIGHT\_IDLE -4.776629e-02
* Sum of Boost\_Pressure\_Left 1.777296e+00
* Sum of Boost\_Pressure\_Left\_Voltage -1.798943e+01
* Sum of Boost\_Pressure\_Right 2.072957e-01
* Sum of Boost\_Pressure\_Right\_Voltage -2.335321e+00
* Sum of Eng\_Hours 1.186036e-01
* Sum of Engine\_RPM -1.117901e-03
* Sum of Engine\_Temp -2.628891e-02
* Sum of Expr1 -3.375233e-02
* Count of FileDetailId 8.789011e-13
* Sum of G\_Force\_Lat -3.064239e+00
* Sum of G\_Force\_Long 1.674250e+01
* Sum of G\_Force\_Vert -1.159652e+01
* Sum of GPS\_Altitude -6.400006e-03
* Sum of GPS\_Heading -2.866604e-04
* Sum of GPS\_Sats\_Used 5.237260e-02
* Sum of GPS\_Time -6.563772e-07
* Sum of Traction\_Right -3.572807e-03
* Sum of Traction\_Left 2.776020e-02
* Sum of Traction\_Diff\_radio 9.445324e-02
* Sum of Scroll\_Speed\_Right 2.513585e-02
* Sum of Scroll\_Speed\_Left 2.937356e-02
* Sum of SCROLL\_REVERSE\_TABLE\_RIGHT 1.571364e-02
* Sum of SCROLL\_REVERSE\_TABLE\_LEFT -2.298484e-02
* Sum of SCROLL\_FORWARD\_TABLE\_RIGHT 1.228333e-02
* Sum of SCROLL\_FORWARD\_TABLE\_LEFT -1.827292e-03
* Sum of RowIndex -6.818553e-07
* Count of MeasurementId -2.664535e-14
* Sum of Main\_Hydraulic\_Pressure\_Right 2.857141e-03
* Sum of Main\_Hydraulic\_Pressure\_Left 1.329919e-02
* Sum of JOY\_STICK\_RIGHT\_REVERSE -5.215495e-02
* Sum of JOY\_STICK\_RIGHT\_FORWARD 5.268926e-02
* Sum of JOY\_STICK\_LEFT\_REVERSE -8.734277e-03
* Sum of JOY\_STICK\_LEFT\_FORWARD 7.848722e-04

From that data the input data with high coefficient are

* **Sum of Boost\_Pressure\_Left**: 1.777296e+00
* **Sum of Boost\_Pressure\_Left\_Voltage**: -1.798943e+01
* **Sum of G\_Force\_Long**: 1.674250e+01
* **Sum of G\_Force\_Lat**: -3.064239e+00
* **Sum of G\_Force\_Vert**: -1.159652e+01
* **Sum of Engine\_Temp**: -2.628891e-02 (this is relatively high, suggesting it has some predictive power)
* **Sum of Traction\_Left**: 2.776020e-02
* **Sum of Traction\_Diff\_radio**: 9.445324e-02
* **Sum of SCROLL\_REVERSE\_TABLE\_RIGHT**: 1.571364e-02
* **Sum of Main\_Hydraulic\_Pressure\_Left**: 1.329919e-02

After removing the coefficients that would have to do with speed, traction and individual scrolls (as this would be operator specific likely) I was left with.

* **Sum of Engine\_Temp**: -2.628891e-02 (this is relatively high, suggesting it has some predictive power)
* **Sum of Traction\_Diff\_radio**: 9.445324e-02

## With traction, torque and speed, Engine\_Temp, Sum of Traction\_Diff\_radio as input information.

Cleaned CSV Loaded Successfully

Number of rows in the training data: 217

Linear Regression Model Performance:

Mean Squared Error: 0.2331635126621323

R-squared: 0.8490397947151831

Feature Weights (Coefficients):

Coefficient

Average of Traction\_Average 0.009450

Average of Engine\_Torque 0.067303

Average of GPS\_Speed -0.178691

Sum of Engine\_Temp -0.037101

Sum of Traction\_Diff\_radio -0.325616

It should be noted that traction\_average has a very small coefficient it’s likely that is has a small impact in prediction however Traction\_Diff\_radio seems to have a bigger impact.

## Realized that I had significantly more good run samples, so I tried a resampling technique to even the bias in the dataset.

Number of samples for each 'Mud Condition' label:

Mud Condition

3 102

2 102

0 102

1 102

4 102

Linear Regression Model Performance:

Mean Squared Error: 0.15146142208261817

R-squared: 0.9197197593842443

Feature Weights (Coefficients):

Coefficient

* Average of Traction\_Average -9.174230e-01
* Average of Engine\_Torque 7.031233e-02
* Average of GPS\_Speed -4.308891e-01
* Sum of Air\_Temp\_Manifold -7.499125e-04
* Sum of Ambient\_Air\_Temp -1.606001e-02
* Sum of BEARING\_PRESSURE\_LEFT\_IDLE -1.487244e-01
* Sum of BEARING\_PRESSURE\_RIGHT\_IDLE 4.010375e-02
* Sum of Boost\_Pressure\_Left 1.290949e+00
* Sum of Boost\_Pressure\_Left\_Voltage -1.311328e+01
* Sum of Boost\_Pressure\_Right -1.465677e-01
* Sum of Boost\_Pressure\_Right\_Voltage 1.263770e+00
* Sum of Eng\_Hours 4.016349e-02
* Sum of Engine\_RPM 1.080917e-03
* Sum of Engine\_Temp -3.856509e-02
* Sum of Expr1 -3.541588e-02
* Count of FileDetailId -2.162922e-12
* Sum of G\_Force\_Lat -2.591186e+00
* Sum of G\_Force\_Long 1.339489e+01
* Sum of G\_Force\_Vert -7.013219e+00
* Sum of GPS\_Altitude -1.836021e-02
* Sum of GPS\_Heading 2.182097e-04
* Sum of GPS\_Sats\_Used 6.293640e-02
* Sum of GPS\_Time -3.984252e-07
* Sum of Traction\_Right 4.639507e-01
* Sum of Traction\_Left 4.731947e-01
* Sum of Traction\_Diff\_radio -5.736842e-01
* Sum of Scroll\_Speed\_Right -3.055726e-03
* Sum of Scroll\_Speed\_Left 6.538784e-03
* Sum of SCROLL\_REVERSE\_TABLE\_RIGHT -1.285545e-02
* Sum of SCROLL\_REVERSE\_TABLE\_LEFT 4.053743e-04
* Sum of SCROLL\_FORWARD\_TABLE\_RIGHT 3.255994e-02
* Sum of SCROLL\_FORWARD\_TABLE\_LEFT -6.033051e-03
* Sum of RowIndex 9.609705e-07
* Count of MeasurementId 5.240253e-14
* Sum of Main\_Hydraulic\_Pressure\_Right 5.588310e-04
* Sum of Main\_Hydraulic\_Pressure\_Left 1.055400e-04
* Sum of JOY\_STICK\_RIGHT\_REVERSE 1.323980e-03
* Sum of JOY\_STICK\_RIGHT\_FORWARD -1.687954e-02
* Sum of JOY\_STICK\_LEFT\_REVERSE 6.367300e-02
* Sum of JOY\_STICK\_LEFT\_FORWARD -6.277796e-02

From that data the input data with high coefficient are

* Average of Traction\_Average -9.174230e-01
* Sum of Boost\_Pressure\_Left 1.290949e+00
* Average of Engine\_Torque 7.031233e-02
* Sum of Engine\_Temp -3.856509e-02
* Sum of Traction\_Right 4.639507e-01
* Sum of Traction\_Left 4.731947e-01
* Sum of Traction\_Diff\_radio -5.736842e-01
* Average of GPS\_Speed -4.308891e-01

After removing the coefficients that would have to do with speed, traction and individual scrolls (as this would be operator specific likely) I was left with.

* Average of Traction\_Average -9.174230e-01
* Average of Engine\_Torque 7.031233e-02
* Average of GPS\_Speed -4.308891e-01
* Sum of Engine\_Temp -3.856509e-02
* Sum of Traction\_Diff\_radio -5.736842e-01

## With traction, torque and speed, Engine\_Temp, Sum of Traction\_Diff\_radio as input information.

Number of rows in the training data: 217

Original data size: 217 samples

Resampled data size: 510 samples

Linear Regression Model Performance:

Mean Squared Error: 0.14970009952955451

R-squared: 0.934804474062676

Feature Weights (Coefficients):

* Average of Traction\_Average 0.008396
* Average of Engine\_Torque 0.073174
* Average of GPS\_Speed -0.304995
* Sum of Engine\_Temp -0.030686
* Sum of Traction\_Diff\_radio -0.754720

Note that average traction seems to reduce it coefficient when other inputs are removed.

## Tried on unknow values here the power bi map

