PREDICTION OF MATERIAL REMOVAL RATE (MRR) IN CHEMICAL MECHANICAL POLISHING PROCESS

Abstract

Developing ML models to predict material removal rate in a chemical polishing process using Multiple Regression and Support Vector Regression Algorithms

Adeniji, David O.

doad224@uky.edu

1.0 Introduction/Project Background

Given two sets of data (Data1 & Data2) with 25 independent variables each, we are to predict the polishing removal rate of material from a wafer for each dataset using three (3) different machine learning approach, Multiple Regression with Gradient descent, Support vector regression with Gradient Descent and the inbuilt sklearn.svm Support Vector Regression module. Table 1 below shows the data column structure and description of the variables.

Table 1. Description of data structure and variables

Column Symbol	Column Name	Description
X1	Machine ID	Numeric ID of machine
X2	Machine Data	Numeric ID of wafer ring location in machine
X3	Timestamp	Seconds
X4	Wafer ID	Number representing ID of wafer
X5	Stage	A or B representing a different type of processing stage
X6	Chamber	Chamber in machine for wafer processing
X7	Usage of Backing Film	A usage measure of polish-pad backing film
X8	Usage of Dresser	A usage measure of dresser
X9	Usage of Polishing Table	A usage measure of polishing table
X10	Usage of Dresser Table	A usage measure of dresser table
X11	Pressurized Chamber Pressure	Chamber pressure
X12	Main Outer Air Bag Pressure	Pressure related to wafer placement
X13	Center Air Bag Pressure	Pressure related to wafer placement
X14	Retainer Ring Pressure	Pressure related to wafer placement
X15	Ripple Air Bag Pressure	Pressure related to wafer placement
X16	Usage of Membrane	A usage measure of polishing membrane
X17	Usage of Pressurized Sheet	A usage measure of wafer carrier flexible sheet
X18	Slurry Flow Line A	Flow rate of slurry type A
X19	Slurry Flow Line B	Flow rate of slurry type B
X20	Slurry Flow Line C	Flow rate of slurry type C
X21	Wafer Rotation	Rotation rate of wafer
X22	Stage Rotation	Rotation rate of stage
X23	Head Rotation	Rotation rate of head
X24	Dressing Water Status	Status of dressing water
X25	Edge Air Bag Pressure	Pressure of bag on edge of wafer
Y	***Material Removal Rate***	***Material Removal Rate***

Variables 1,2,3,4,5,6 are unique machine data and would not affect the material removal rate (MRR). Therefore Columns 1 to 6 are dropped, and only Columns 7 to 25 are used.

2.0 Feature/Variable Selection

Before feeding the data variables into a machine learning algorithm, it is important to understand the relationships between the independent and dependent variables, this will help in understanding the variables that would be useful for better model prediction.

2.1 Correlation Plot between Variables

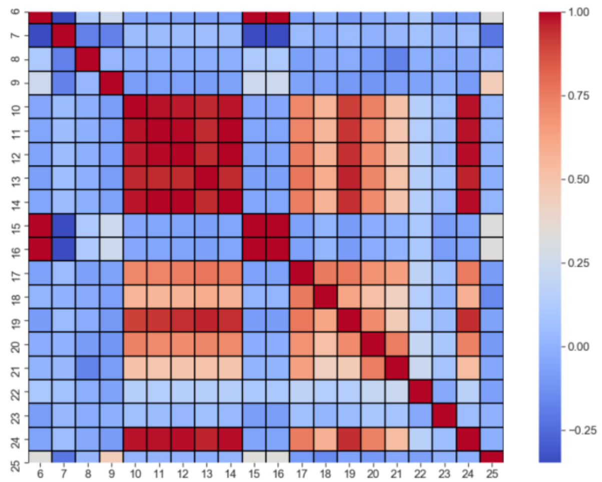
Using the seaborn library in python, the correlation between all the variables was computed. Results showed that only few independent variables are correlated with the MRR, also some independent variables are correlated between themselves. In order to avoid ***Multi-Collinearity,*** it is important to drop some variables that are highly correlated among themselves. Table1 shows the variable correlation with the MRR, while Figure 1a and 1b shows the correlation heatmap among all variables for dataset 1 and 2 respectively.

Table 2. Variable Correlation with the Material Removal Rate for each Dataset

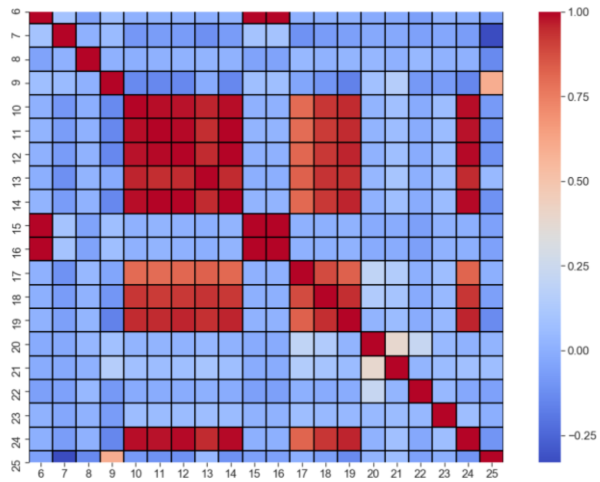
Column Symbol	***Column Name***	***Data 1***	***Data 2***
X7	Usage of Backing Film	0.336	0.053
X8	Usage of Dresser	0.213	0.331
X9	Usage of Polishing Table	0.022	0.137
X10	Usage of Dresser Table	0.449	0.599
X11	Pressurized Chamber Pressure	0.013	0.081
X12	Main Outer Air Bag Pressure	0.014	0.112
X13	Center Air Bag Pressure	0.009	0.109
X14	Retainer Ring Pressure	0.011	0.044
X15	Ripple Air Bag Pressure	0.012	0.111
X16	Usage of Membrane	0.336	0.053
X17	Usage of Pressurized Sheet	0.336	0.053
X18	Slurry Flow Line A	0.089	0.004
X19	Slurry Flow Line B	0.149	0.056
X20	Slurry Flow Line C	0.059	0.132
X21	Wafer Rotation	0.091	0.020
X22	Stage Rotation	0.043	0.066
X23	Head Rotation	0.069	0.052
X24	Dressing Water Status	0.002	0.000
X25	Edge Air Bag Pressure	0.010	0.098
Y	***Material Removal Rate***	1.000	1.000

From Figure 1a, shows that variables 10, 11, 12, 13, 14, 19 and 24 are strongly correlated in Dataset 1, and as discussed some of these variables must be dropped to avoid multi-collinearity. Also, Dataset 2 heatmap shows variables 6, 15 and 16 to be strongly correlated. From Table 2, it is obvious we must include variables 7, 10, 16 or 17, in our model, as they have the strongest correlation with the dependent variable (MRR). While for Dataset 2, we must include variables 8 and 10.

Figure 1. Correlation Heat Map



a. Dataset 1



b. Dataset 2

2.2 Preferred Feature/Variable Selection

Sequel to section 2.1, we then proceed to select best variables to be used in developing our model.

Table 3. Selected Variables across Datasets		
	Selected Variables	
	Column ID	***Column Name***
Dataset 1	X7, X8, X10, X16	Usage of Backing Film, Usage of Dresser, Usage of Dresser Table, Usage of Membrane
Dataset 2	X8, X10	Usage of Dresser, Usage of Dresser Table.

3.0 Data Pre-processing

3.1 Mean Normalization

The original dataset variables are in different value ranges, however, in order to get better prediction capability, there is need for the variables to be in the same scale. Both dataset 1 and 2 datapoints were first normalized to a value range of -1 and 1, using the ***Mean Normalization Approach,*** whose equation is shown in Eq. 1.:

$$x' = \frac{X - X_{average}}{X_{max} - X_{min}} \qquad (1)$$

3.2 Kernel Transformation

For the SVR Gradient Descent approach, the data must first be processed using Kernel transformation. The kernel functions transform the data into a higher dimensional feature space to make it possible to perform the linear separation.

3.3 Train_Test_Split

For model validation and accurate training, we must split the data into Train and Test portions, where the train data is first used to develop the model, and the test data is used for validation. For all datasets and approach, only 70% of the data is used for Training, while 30% is used for testing.

4.0 ML Algorithm Parameter and Results

For this case study, we would only compare three ML algorithms (Multiple Regression with Gradient descent, Support vector regression with Gradient Descent and the inbuilt sklearn.svm Support Vector Regression module)

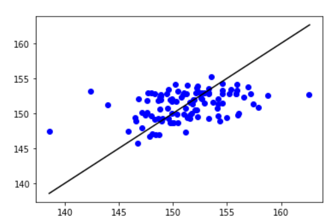
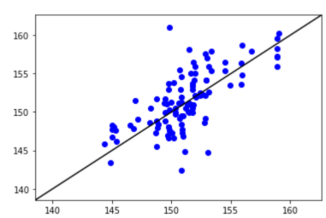
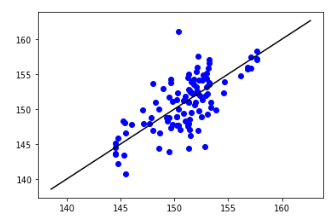
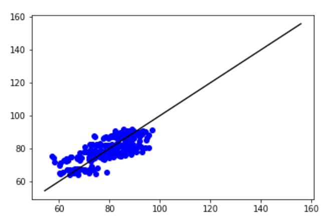
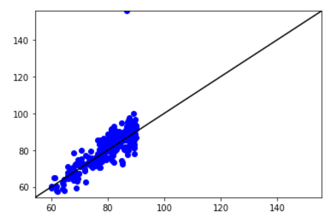
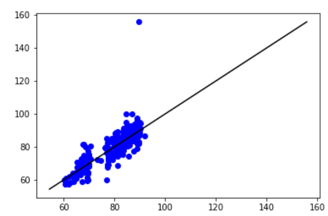
Table 4a. Result Output for Dataset 1				
		Multiple Regression with Gradient descent	***Support vector regression with Gradient Descent***	***sklearn.svm Support Vector Regression module***
Best Variables Combination		X7, X8, X10, X17, X19, X20	X8, X10	X8, X10
Best R-Squared Value		0.16	0.45	**0.48 *Best***
Best RMSE Value		3.24	2.87	
Prediction vs Target Plots				
Gradient Descent Parameters	alpha	0.1	0.1	----
	iteration	20,000	50,000	----
	lamda	----	0.001	----
	Gamma	----	100	$1/n_{features}$
	C	----	----	100
	Epsilon	----	----	4
	Kernel	----	rbf	rbf

Table 4b. Result Output for Dataset 2

		Multiple Regression with Gradient descent	***Support vector regression with Gradient Descent***	***sklearn.svm Support Vector Regression module***
Best Variables Combination		X7, X8, X9, X10, X17, X19, X20	X8, X10	X7, X8, X9, X10
Best R-Squared Value		0.51	0.63	**0.71 *Best***
Best RMSE Value		6.18	6.43	13.5
Prediction vs Target Plots				
Gradient Descent Parameters	alpha	0.1	0.1	----
	iteration	20,000	50,000	----
	lamda	----	0.001	----
	Gamma	----	30	$1/n_{features}$
	C	----	----	110
	Epsilon	----	----	1
	Kernel	----	rbf	rbf

5.0 Observations & Conclusion

5.1 Observations

While working on this project, the following observations were made across the methods:

Multiple Regression Model

1. While the correlation plots suggested that we should use only two variables, the model accuracy while using the four variables, X7, X8, X10, X16, were significantly low, however, the accuracy increased when other variables were added.
2. The iteration value (length of model training) significantly influenced the model accuracy.
3. There appears to be a limit for the alpha value (about 0.9), beyond which the result becomes invalid.

Support Vector Regression with Gradient Descent

4. As suggested by the correlation plots, variables X8, and X10 gave the best model accuracy, while the addition of other models seemed to reduce the accuracy.
5. Selecting the optimal gamma value was quite difficult, however, while trying different values, a gamma value of 10 appear to be the best.
6. There appears to be a limit for alpha and lambda values (about 0.9), beyond which the result becomes invalid.

Support Vector Regression in sk.learn library

7. As suggested by the correlation plots, variables X7, X8, X9, and X10 gave the best model accuracy of 71% R-square value, while the addition of other models seemed to reduce the accuracy.
8. Selecting the optimal value for the ***regularization parameter C and epsilon*** was quite difficult, however, I wrote a code to find the optimal value for both parameters and that seemed to fix it.

Overall

9. Depending on the random seed number used in the code, the model accuracy changes significantly
10. This process improved my coding skills, as I had to try different approach, just to get the best model accuracy

5.2 Conclusion

1. Overall, looking at Table 4a and 4b above, it is shown that variables X7, X8, X9, and X10, which are Usage of *Backing Film, Dresser, Polishing Table and Dresser Table* respectively are all important across the three models. This is expected, because ***the longer the Usage of Backing Film, Dresser, Polishing Table and Dresser Table, the lesser the material removal rate.***
2. Both Support Vector Regression method performed much better than the Multiple regression approach. This is because the SVR approach is more mathematically robust than the Multiple Regression and can also handle non-linear data perfectly well.
3. Reflecting on the whole process, selecting the optimal variables and parameter for the model appear to be much difficult than expected. It is thereby important to properly work on the feature selection step and parameter tuning in order to get the best model accuracy.
4. We can safely conclude that only 4 variables, namely X7, X8, X9, and X10, (Usage of *Backing Film, Dresser, Polishing Table and Dresser Table* respectively) are much relevant to predicting the material removal rate.

6.0 Appendix

Additional Figures used for Feature Selection