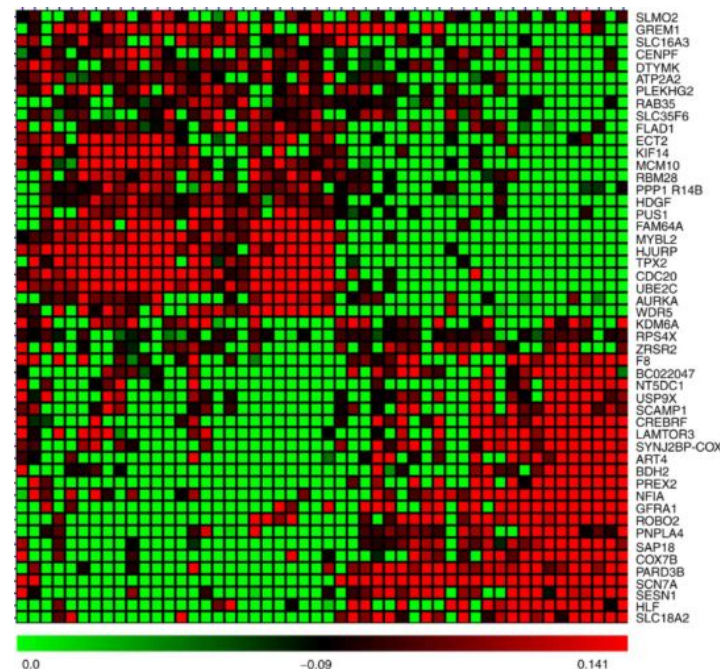
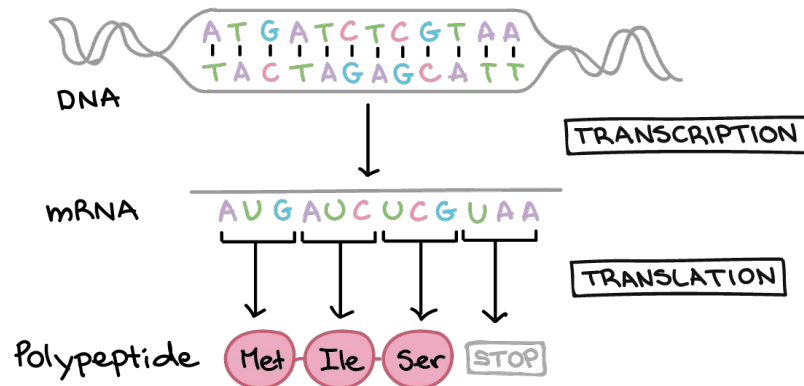


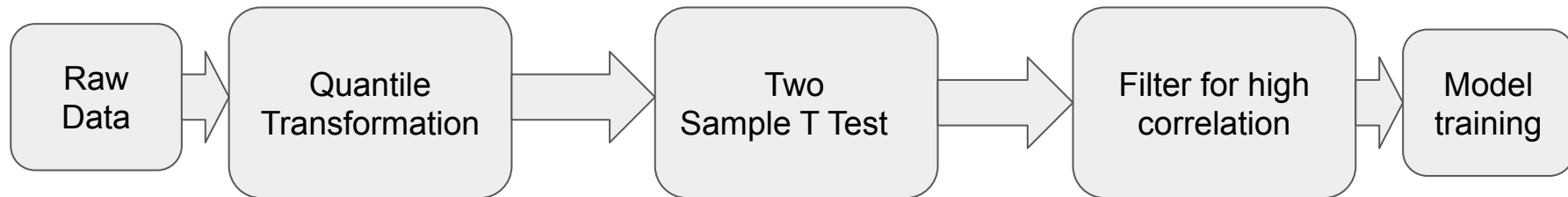
Mid Project Bootcamp

Predicting if a Patient has Cancer from Gene Expression Set

Introduction to Gene Profiling

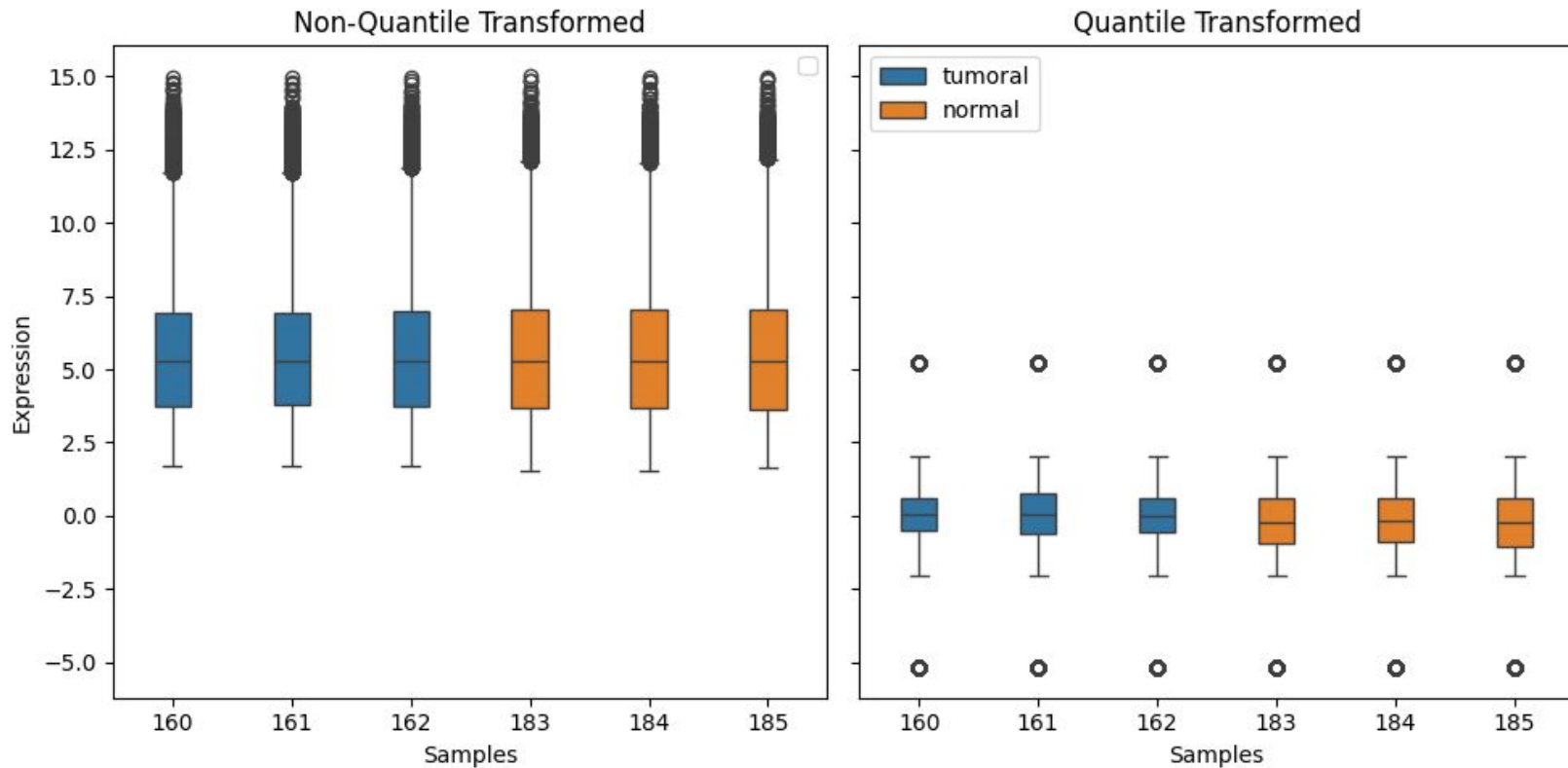


4 steps to clean the dataset before model training



Quantile Transformation

Quantile transformation is applied to gene expression data to normalize its distribution

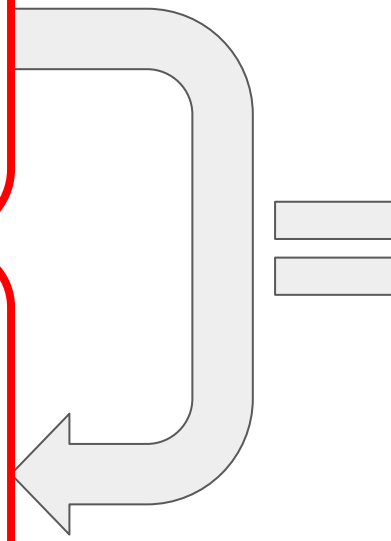


Two Sample T Test

A two-sample t-test is employed to assess whether there is a significant difference in the mean expression levels between tumoral and normal genes.

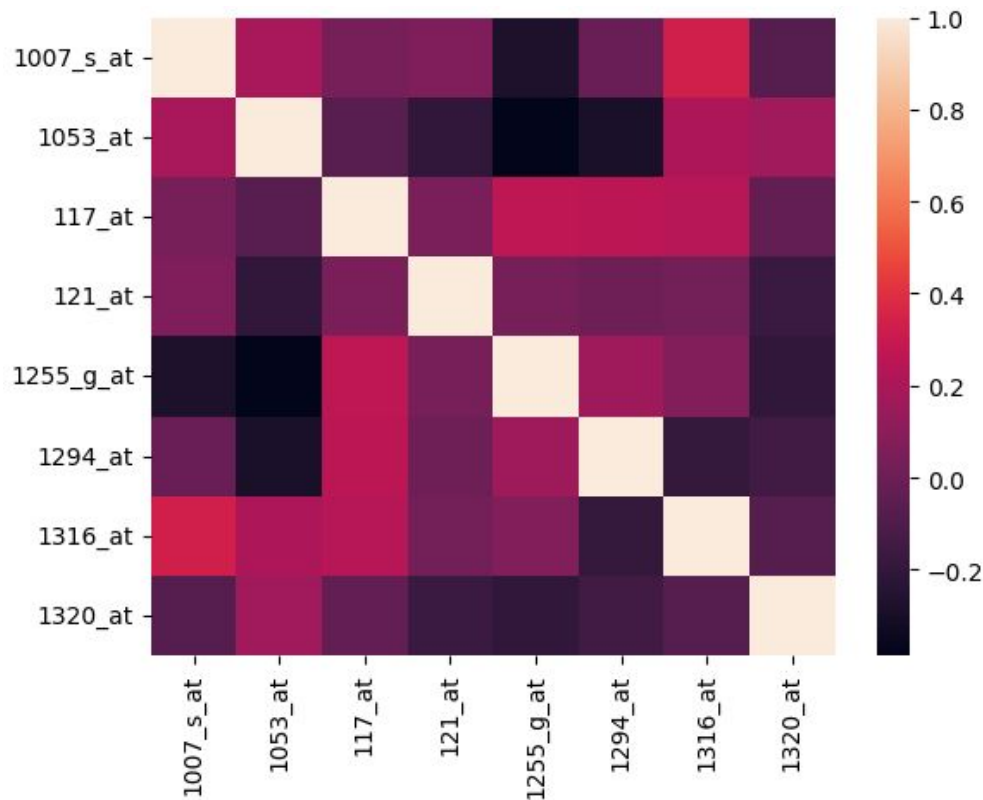
samples			type	NM_004900	AA085955
0	GSM563920_1	primary_breast_cancer		0.000000	0.618521
1	GSM563922_3	primary_breast_cancer		-5.199338	1.711675
2	GSM563923_4	primary_breast_cancer		-0.910955	-1.359737
3	GSM563924_5	primary_breast_cancer		0.732656	-0.857254
4	GSM563925_6	primary_breast_cancer		-0.018165	0.090945

samples			type	NM_004900	AA085955
134	GSM564101_182	normal		-0.201152	0.553808
135	GSM564102_183	normal		-0.938814	-1.027154
136	GSM564103_184	normal		-0.857254	-1.512390
137	GSM564104_185	normal		-1.457684	0.805918
138	GSM564105_186	normal		-0.532755	-0.575109



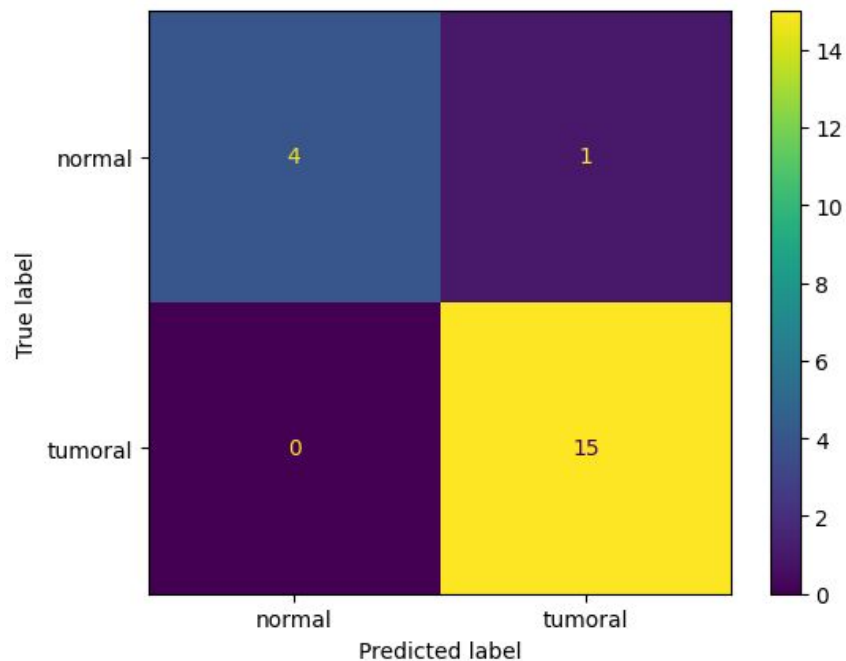
Filter for Multi correlation coefficients between Genes

Filtering genes with high correlation is a crucial step to reduce overfitting in the final model.

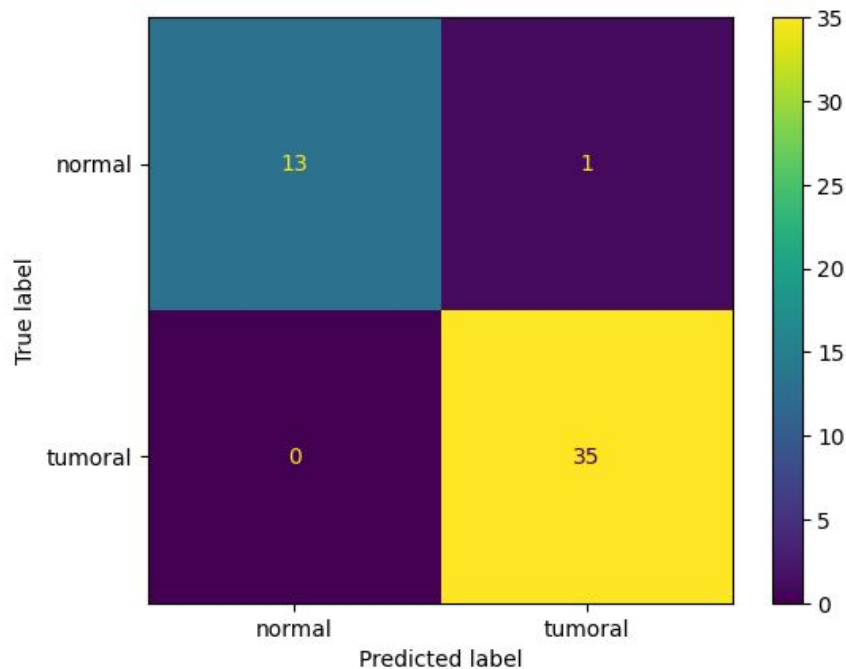


Logistic Model for Predicting Prostate Cancer

A logistic regression model was employed to predict whether a sample belongs to the normal or tumoral category based on the filtered gene set.



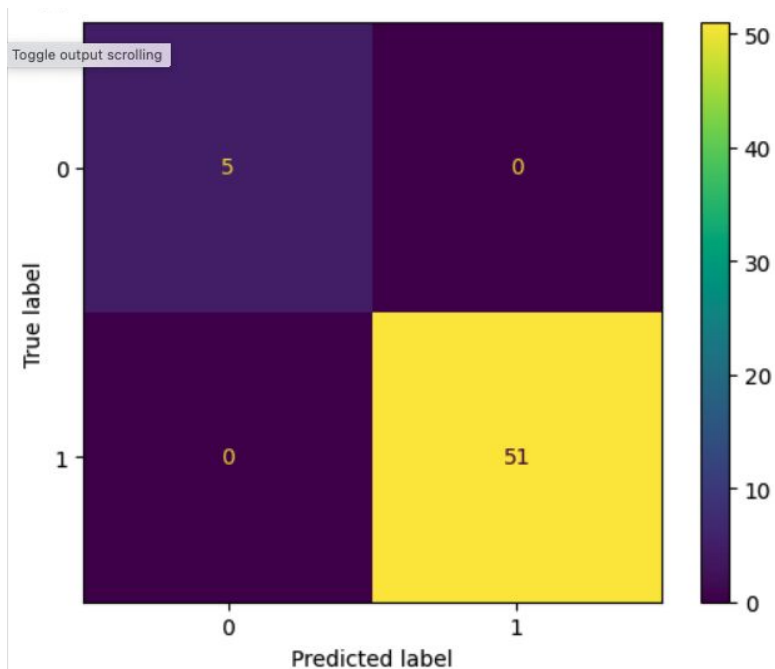
Test data set



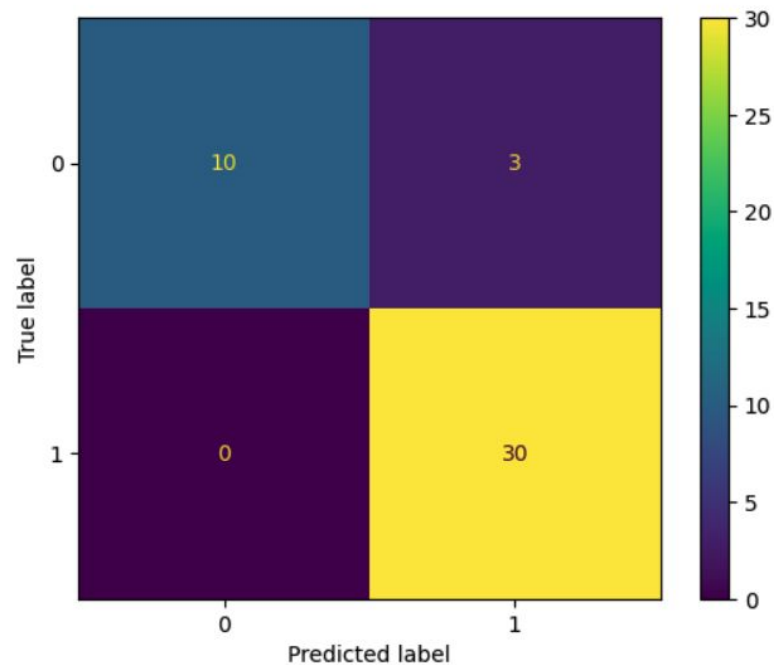
Fresh new Data set

Logistic Model for Predicting Breast Cancer

A logistic regression model was employed to predict whether a sample belongs to the normal or tumoral category based on the filtered gene set.



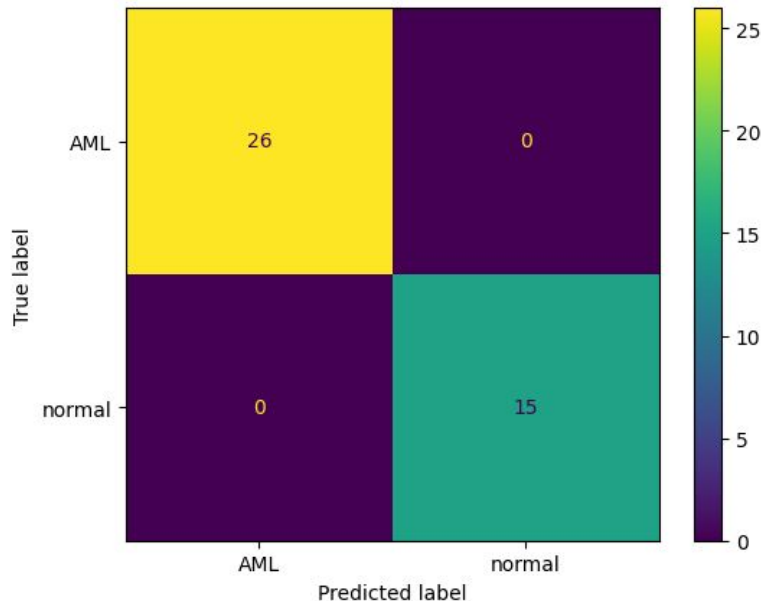
Test data set



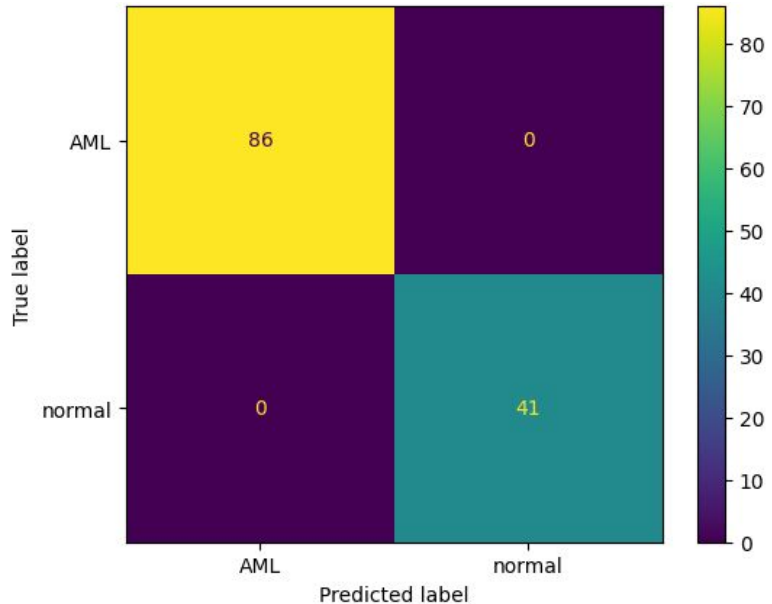
Fresh new Data set

Logistic Model for Predicting Leukemia AML-type Cancer

A logistic regression model was employed to predict whether a sample belongs to the normal or tumoral category based on the filtered gene set.



Test data set



Fresh new Data set