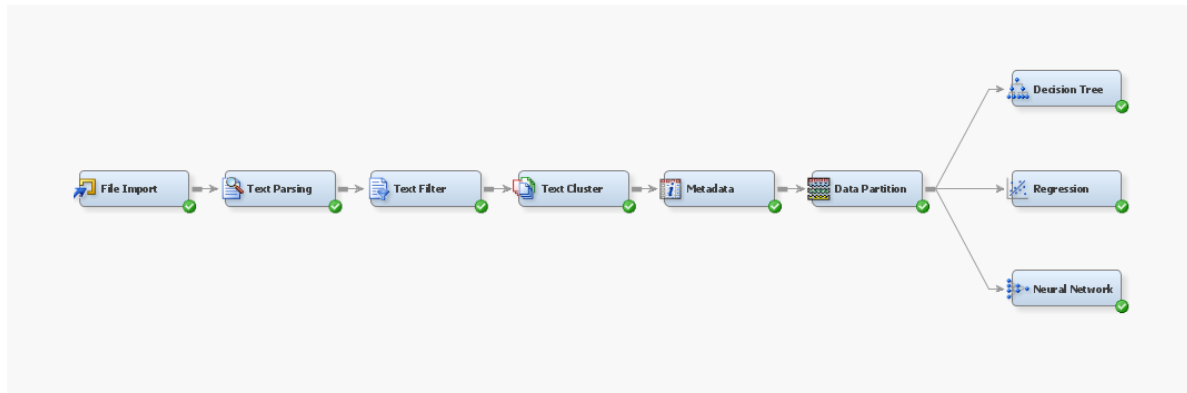# Sentiment Assignment

The best model to predict Sentiment is **Neural Network** with Average Squared Error (Train 0.17, Validation 0.19) Misclassification Rate (Train 0.26, Validation 0.29).

1- File Import
The "Sentiment Analysis Dataset" contains 1,048,588 tweets and three variables ItemID, Sentiment, SentimentText. ItemID is irrelevant to the analysis, and it was dropped. Maximum rows to import option was set to 10,000.

2- Text Parsing
"Different parts of speech" option is set to "No".

3- Text Filter
Term Weight by IDF and Frequency Weighting is set to log.

4- Text Cluster
SVD Resolutions is set to High, and number of clusters is 40.

Descriptive Terms
+last +night +back fun dont +leave +bed +know +school +year +find +time +laker +hour +show ...
followfriday It +love +good gt +follow +back +'good day' sentiment140 +day actress +secret gorgeous +mention tweeps ...
+time im +sad lol +face +today +feel +know +work +miss +sick +want +great +cry +mom ...
+work +hate +look +leave +job +home +bear +goodbye +fact +stick +holiday +hand +watch +tomorrow gt ...
+good +love +day It followfriday +song +nice +morning +want +work +hour +lovely awesome +late +mean ...
myweakness iremember +girl +wait +lose wanna iphone squarespace +hope +win musicmonday cant +morning +life +find ...
+happy +wish +tonight +home +watch +people +well +bad +right +thing +trend +topic +'trending topic' shaundiviney +tomorrow ...
inaperfectworld +help +hurt haveyouever +rain +break +start dontyouhate +die +suck down +head +big awesome +hear ...
+miss +want +day +today +know +feel im +sad lol +time +great +follow +night +good +love ...

5- Metadata

Name	Hidden	Hide	Role	New Role	Level	New Level
ItemID	Y	Default	Rejected	Rejected	Interval	Default
Sentiment	N	Default	Input	Target	Interval	Binary
SentimentText	N	Default	Text	Rejected	Nominal	Default
TextCluster_SVD1	N	Default	Input	Default	Interval	Default
TextCluster_SVD10	N	Default	Input	Default	Interval	Default
TextCluster_SVD100	N	Default	Input	Default	Interval	Default

6- Data Partition
Training 70 Validation 30

7- Models

    a.   Decision Tree

        Average Squared Error (Train 0.197877, Validation 0.21347)

        Misclassification Rate (Train 0.297228, Validation 0.327115)

Fit Statistics	Statistics Label	Train	Validation
NOBS	Sum of Frequencies	6998	3002
MISC	Misclassification Rate	0.297228	0.327115
MAX	Maximum Absolute Error	0.940063	0.940063
SSE	Sum of Squared Errors	2769.491	1281.675
ASE	Average Squared Error	0.197877	0.21347
RASE	Root Average Squared Error	0.444834	0.462028
DIV	Divisor for ASE	13996	6004
DFT	Total Degrees of Freedom	6998	.

    b.   Regression

        Average Squared Error (Train 0.176966, Validation 0.186725)

        Misclassification Rate (Train 0.266505, Validation 0.285809)

Fit Statistics	Statistics Label	Train	Validation
AIC	Akaike's Information Criterion	7586.155	.
ASE	Average Squared Error	0.176966	0.186725
AVERR	Average Error Function	0.52759	0.549134
DFE	Degrees of Freedom for Error	6897	.
DFM	Model Degrees of Freedom	101	.
DFT	Total Degrees of Freedom	6998	.
DIV	Divisor for ASE	13996	6004
ERR	Error Function	7384.155	3297.002
FPE	Final Prediction Error	0.182149	.
MAX	Maximum Absolute Error	0.983206	0.971254
MSE	Mean Square Error	0.179557	0.186725
NOBS	Sum of Frequencies	6998	3002
_NW_	Number of Estimate Weights	101	.
RASE	Root Average Sum of Squares	0.420673	0.432117
RFPE	Root Final Prediction Error	0.426789	.
RMSE	Root Mean Squared Error	0.423742	0.432117
SBC	Schwarz's Bayesian Criterion	8278.346	.
SSE	Sum of Squared Errors	2476.81	1121.099
SUMW	Sum of Case Weights Times Freq	13996	6004
MISC	Misclassification Rate	0.266505	0.285809

    c.   Neural Network

        Average Squared Error (Train 0.17202, Validation 0.18913)

        Misclassification Rate (Train 0.257931, Validation 0.289474)

Fit Statistics	Statistics Label	Train	Validation
DFT	Total Degrees of Freedom	6998	.
DFE	Degrees of Freedom for Error	6691	.
DFM	Model Degrees of Freedom	307	.
_NW_	Number of Estimated Weights	307	.
AIC	Akaike's Information Criterion	7838.958	.
SBC	Schwarz's Bayesian Criterion	9942.946	.
ASE	Average Squared Error	0.17202	0.18913
MAX	Maximum Absolute Error	0.9636	0.962766
DIV	Divisor for ASE	13996	6004
NOBS	Sum of Frequencies	6998	3002
RASE	Root Average Squared Error	0.414754	0.434891
SSE	Sum of Squared Errors	2407.599	1135.539
SUMW	Sum of Case Weights Times Freq	13996	6004
FPE	Final Prediction Error	0.187806	.
MSE	Mean Squared Error	0.179913	0.18913
RFPE	Root Final Prediction Error	0.433366	.
RMSE	Root Mean Squared Error	0.424162	0.434891
AVERR	Average Error Function	0.516216	0.559305
ERR	Error Function	7224.958	3358.067
MISC	Misclassification Rate	0.257931	0.289474
WRONG	Number of Wrong Classifications	1805	869