

# Analyze the Sales Report of a Clothes Manufacturing Outlet

Dated: 19<sup>th</sup> April 2021

Developer: Puneet Kumar Virmani

## Solutions:

1. To automate the process of recommendations, the store needs to analyze the given attributes of the product, like style, season, etc., and come up with a model to predict the recommendation of products (in binary output – 0 or 1) accordingly.

The Attribute DataSet.csv file contains all the attributes and the recommendations for each dress in binary. For binary output models, logistic regression can be used to build a suitable model to recommend products (as mentioned earlier, a value of 1 denotes a positive recommendation and 0 denotes a negative recommendation). To perform logistic regression on a given dataset, we need to decide two major attributes of the model – the dependent and independent variables. The required values are: Dependent variable: Recommendation Independent Variables: All other variables, except the Dress\_ID, since it is only an identifier. Outlier treatments: None

## Code:

| Model               | Model Code  | Model Accuracy |
|---------------------|---|----------------|
| Logistic Regression | <pre>LRModel &lt;-<br/>bayesglm(Recommendation ~ Style +<br/>Price + Rating + Size + Season +<br/>NeckLine + SleeveLength +<br/>waiseline + Material + FabricType +<br/>Decoration + `Pattern Type`<br/>+ Avg_sales_P_ID<br/>, data = train_data, family=<br/>binomial)</pre> | 63%            |
| Native Bayes        | <pre>classifier_cl &lt;-<br/>naiveBayes(Recommendation ~ .,<br/>data = train_data)</pre>  | 67%            |
| KNN                 | <pre>prc_test_pred &lt;- knn(train =<br/>train_data, test = test_data, cl =<br/>train_data\$Recommendation, k=5)</pre>  | 67%            |
| SVM                 | <pre>svmfit = svm(Recommendation ~ .,<br/>data = Att_Data_set, kernel =<br/>"linear", type = 'C-classification')</pre>  | 96%            |

## Result:

From the significance codes for each attribute, we can see that season and Average sale price make an impact on the recommendation, both positively affecting the recommendation. Other than that, we can see that the increased number of factors and comparatively lesser number of entries make the predictions slightly difficult.

However, the residual deviance is lower than the null deviance, which implies that using the independent variables makes it closer to predicting the actual values of recommendation.

With the given model, the new data or attributes can be fed into the model to get recommendations. We have run our data on 4 different concepts to see the output and SVM is only model with the higher accuracy.