UNIVERSITÄT HOHENHEIM

Demand Forecasting Report: How To Predict Your Promotion

CONSULTANTS

Jaron Lins
Julius Keune
Leonard Hegemann
Yaro Dusel

I. ABNORMALITIES IN YOUR DATA

The biggest challenge in your data was uniquely identifying your products. By examining different entries, one can see that the same product identifier corresponds to different categories depending on the store. **We recommend relabelling your products with a single, unique identifier across all stores**. This change will simplify future analyses to optimize your forecasts even more. Our analysis also shows that you order approximately 25,000 more units than you sell. **To reduce storage costs, consider restocking in a more sustainable manner.** However, if restocking is automated, this discrepancy might indicate potential theft which should be monitored.

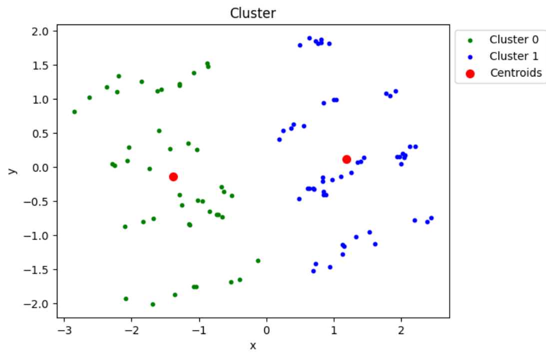

Fig. 1: Scatter Plot showing the clusters for the best k = 2

II. PATTERNS IN YOUR DATA

By clustering your data into an optimal number of groups using measures like the silhouette score, meaningful patterns emerge. Figure 1 shows a visualization of the two clusters. For more detail, please refer to the bar charts in Figures 2 and 3. From Figure 2, we see that the **clusters differ primarily in product price**. Figure 3 confirms that each cluster contains distinct categories: the higher-priced Cluster 0 (blue) includes Clothing, Electronics, and Furniture, while the lower-priced Cluster 1 (orange) includes Groceries and Toys. These findings indicate that your product range can be divided into **two meta-categories: non–fast-moving goods in Cluster 0 and fast-moving goods in Cluster 1**.

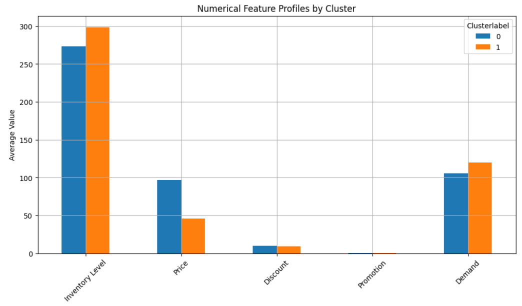

Fig. 2: Bar Chart showing the mean values of the numerical attributes of each cluster

III. TREE BASED DECISIONS

For the task of predicting whether to promote a product, it is important to focus only on the most relevant variables and exclude any tracked data that do not contribute to this decision. Figure 4 highlights the key predictors used in our decision-tree analysis. **First, any product already discounted by more than 10 percent should always be promoted; conversely, products with discounts below 7.5 percent should never be promoted. For products whose discounts fall between 7.5 percent and 10 percent, promotion should depend on the expected demand: if the expected demand is below 129.5 units, the product should not be promoted, but if the demand meets or exceeds 129.5 units and the weather is not sunny (<=1.5), the product should be promoted**. By following these clear, data-driven guidelines, you can make consistent and reliable promotion decisions for each product.

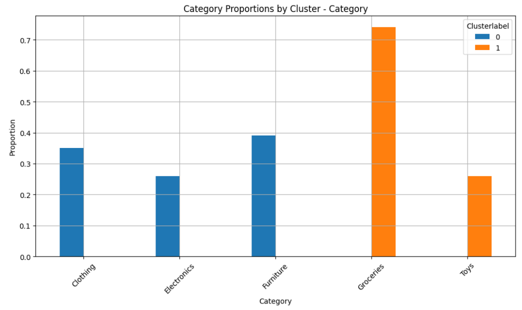

Fig. 3: Bar Chart showing the different categories by their cluster

IV. CRITICAL ASSESSMENT

After giving these clear recommendations for action, we need to assess our model critically. Our decision tree isn't the highest-performing model—**a deeper tree would yield better accuracy at the expense of loosing the interpretability**. Furthermore, our predictions are based on a small sample, so **accuracy could improve with more data.** Another limitation is that you can't update the model on the fly: **any new data requires full retraining**. The tree is also **sensitive to small changes**, since each split depends on the previous one. If a small change causes the root node to split on a different feature or threshold, that difference will flow down through every branch and alter the entire tree. Despite the disadvantages our decision tree **captures different types of non-linear relationships**, delivers **solid predictions, remains easy to interpret, and is very cost-efficient to compute**.

Decision Tree Visualization (Gini)

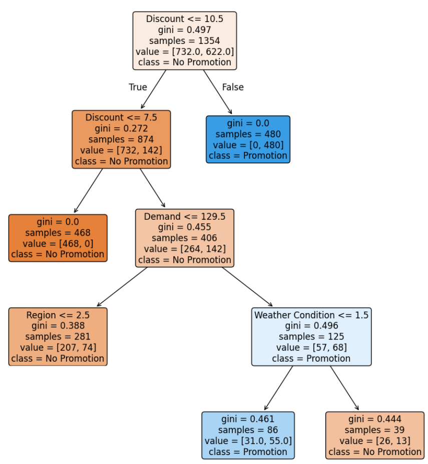

Fig. 4: Decision Tree to predict whether a product gets promoted

Jaron.lins@uni-hohenheim.de | Julius.keune@uni-hohenheim.de | Leonard.hegemann@uni-hohenheim.de | Yaro.dusel@uni-hohenheim.de