**The National Champion Analytics Case Competition**

*Retail Champion: A Predictive Analysis of New Product Demand*

University of Massachusetts Lowell

Team Members: Modupe Ajala, Michael Wile, Dana Lantion

Faculty Advisor: Dr. Amit Deokar

**Executive Summary**

The Champion Brand is a subsidiary of Hanesbrands which was spun off by the Sara Lee Corporation in 2006. Champion is an American manufacturer of clothing that specializes in sportswear. The company plans on launching 20 new fashion products in March 2020 at its Brick & Mortar stores and online. The Champion team wants to know the 13-week projected demand for these products and whether the product will succeed or fail and how soon success or failure can be spotted in order to make plans accordingly. They also want to know if the products will perform differently across online and brick & mortar channels. Our analysis answered their above-mentioned questions using the following steps explained below.

**Data Understanding**

We were provided weekly historical data for 1000 ongoing products. We observed from the historical data that it contained similar features to the 20 new products Champion team plans on launching. A data dictionary was also provided, listing what each field in the dataset represented. There were 24 variables in the historical data representing each product in their lowest-highest level of identification. Finally, we were given a new product dataset that contained all the new products and their styles which we then mapped to get the best data point representation in the historical data.

**Data Preprocessing**

Upon reviewing the data, we noticed that though it had no missing values, which was great, there was extreme variance in 2017 product demand compared to that of the years 2018 and 2019, so we excluded 2017 data in our analysis. We also noticed there was no data for November and December 2019, so we selected months from January 2018 to June 2018 and then from January 2019 to June 2019. We treated each 6-month period as a whole season. Both six-month periods were then concatenated. We selected historical data that best represented new products. The historical data was selected based on *style group*, *color*, and *fabrication*. We created datasets that had as many observations as possible while being specific enough to address each new product. If there was more than one observation for a time period, we averaged the demand for those observations.

**Modeling**

We used 2018 and 2019 data for our analysis since demand changed so drastically from 2017. We split the data into train and test sets; the first 39 weeks data was used for training and the later 13 weeks data was used for testing. We imputed missing values with a method called *seasonal adjustment with linear interpolation*. We transformed values by differencing them by 1 period and rescaled values so both the minimum and maximum values were within a –1 to 1 boundary. For this competition, we chose to run unmodified models to simulate quick turnover time series simulations needed for time sensitive supply chain logistics. This will provide comparative results for all products given a specific model. If time isn’t a constraint, it would be ideal to customize the time series models per product and then reevaluate train/test performance for each product.

We employed three different models:

1. **Seasonal Naive**: This method is like the naive method but predicts the last observed value of the same season of the year. This method works for highly seasonal data like ours.
2. **ARIMA**: ARIMA stands for **Autoregressive Integrated Moving Average model**. It is a forecasting technique that projects the future values of a series based entirely on its own inertia. Its main application is in the area of short-term forecasting.
3. **Long short-term memory (LSTM):** This is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series.

**Evaluation**

Our LSTM model performed very well overall. Generally, we got an average RMSE of around 4 in our test/train set. Some were higher and some were lower than that. LSTM generally improved prediction by more than double, on average, in comparison to the ARIMA model. ARIMA performed alright with some products but there was a lot of room for improvement. On average, we saw a slight improvement of around 1 to 2 RMSE from the Seasonal Naive model to the ARIMA. Seasonal Naive model performed the worst out of the group as expected but it provided a baseline to improve our other models.

**Deployment/ Recommendations**

LSTM model was our best performing algorithm, we are confident that if given more time and data, this algorithm could be fine-tuned and improved even further to include: (a) more layers, (b) more epochs, and (c) different optimization methods. We make the following recommendations based on our LSTM model's performance.

If both train/test set are accurate as desired and the respective forecasting model repeatedly predicts a demand below zero, we would not recommend the product. If the model consistently forecasts a demand that is continuous with past performance or increases with desirable train/test results, we recommend the product. Certain forecasts of products could be improved with LSTM models adjustment, but we think it is important to have a single model that allows us compare performance to each new product across the board to simulate time sensitive supply chain logistics.

|  |  |
| --- | --- |
| **Products Recommended** | |
| **Brick and Mortar** | **Online** |
| STYLE009\_BLUE\_------  STYLE009\_PINK\_------  STYLE329\_BLACK\_40GEAR  STYLE366\_BLACK\_------  STYLE179\_BLUE\_------ | STYLE031\_BLACK\_407Q88  STYLE082\_WHITE\_  STYLE087\_GREY\_  STYLE104\_BLUE\_Y06085  STYLE105\_PURPLE\_Y06145  STYLE189\_BLACK\_------  STYLE111\_RED\_Y06145 |

|  |  |
| --- | --- |
| **Products Not Recommended** | |
| **Brick and Mortar** | **Online** |
| STYLE042\_GREY\_549310  STYLE203\_GREY\_549314  STYLE264\_RED\_------  STYLE276\_BLACK\_549333  STYLE359\_BLUE\_------ | STYLE116\_BLACK\_549314  STYLE207\_BLUE\_407D55  STYLE357\_GREY\_------ |