Predicting Fashion Product Success:

A Machine Learning Approach to Trend Score Forecasting

Project AURORA Technical Report

Table of Contents

Abstract

Fashion retail operates under conditions of profound uncertainty: trends emerge and fade unpredictably, consumer preferences shift with seasons and social influences, and the gap between production decisions and consumer feedback can span months. Retailers must commit to inventory and marketing investments with limited information about which products will resonate with customers. This work investigates whether machine learning can help navigate this uncertainty by predicting trend scores for fashion products based on their attributes and early engagement signals. Using data from 12,000 fashion products across seven categories, we develop a gradient boosting model that predicts trend scores with R-squared of 0.78 and mean absolute error of 9.2 points on a 100-point scale. Feature analysis reveals that engagement metrics, particularly page views and wishlist additions, drive predictions, but product attributes like price point, discount level, and seasonal alignment also contribute meaningfully. We discuss applications to inventory planning, pricing optimisation, and marketing resource allocation, while acknowledging limitations around causality and the self-fulfilling nature of trend predictions.

1. Introduction

The fashion industry presents a fascinating forecasting challenge. Unlike many retail categories where demand is relatively stable and predictable, fashion demand is driven by trends that emerge, peak, and decline on timescales ranging from weeks to seasons. A dress style that sells out in spring may languish on clearance racks by autumn. A colour that dominates one season may feel dated the next. Predicting which products will succeed requires understanding not just consumer preferences but their evolution over time.

Traditional fashion forecasting relies heavily on human expertise. Buyers attend fashion weeks, analyse street style photography, and synthesise their observations into purchasing decisions made months before products reach consumers. This process has worked well enough to sustain the industry, but it has obvious limitations: human attention is finite, expertise is concentrated in a few individuals, and systematic biases inevitably creep in. Promising trends may be overlooked; past successes may be over-weighted.

Machine learning offers a complementary approach. Rather than replacing human judgment, it can process signals at scale that humans cannot easily synthesise: thousands of products, millions of customer interactions, patterns across categories and time periods. The question is whether these signals contain predictive information that can be extracted and acted upon. Early engagement data, the clicks and wishlist additions that accumulate in a product's first days online, might reveal consumer interest before sales data becomes available. Product attributes themselves might carry predictive power: certain colours, materials, or price points may systematically perform better in certain contexts.

This project investigates these possibilities using data from a fashion retail platform. We develop a model that predicts trend scores, a composite metric reflecting sales velocity, customer engagement, and product popularity, from product attributes and early engagement signals. The results suggest meaningful predictability, though with important caveats about causality and the limits of forecasting in a domain shaped by social dynamics.

2. Related Work

2.1 Fashion Demand Forecasting

Fashion demand forecasting has attracted substantial research attention given its commercial importance. Traditional approaches use time series methods adapted to fashion's seasonal patterns (Thomassey, 2010). More recent work has incorporated external signals including social media trends, search volume, and weather patterns to improve forecast accuracy (Ren et al., 2017). The common challenge across methods is handling the short product lifecycles and limited historical data characteristic of fashion retail.

A parallel stream of research focuses on trend detection rather than demand forecasting. These approaches identify emerging trends from social media, runway coverage, or e-commerce patterns (Al-Halah et al., 2017). While distinct from demand forecasting, trend detection can inform inventory and marketing decisions by identifying which styles are gaining cultural traction.

2.2 Engagement-Based Prediction

The use of early engagement signals to predict product success is well-established in digital contexts. Netflix uses viewing patterns to predict which shows will succeed; Amazon uses browsing behaviour to optimise recommendations. In fashion e-commerce, Wu et al. (2019) demonstrated that wishlist additions predict subsequent purchases, while Chen et al. (2020) showed that page view velocity in a product's first week correlates with eventual sales volume. Our work extends this literature by combining engagement signals with product attributes in a unified predictive model.

3. Data and Feature Engineering

Our dataset comprises 12,000 fashion products from a mid-market retail platform, spanning seven categories: Dresses, Tops, Bottoms, Outerwear, Shoes, Bags, and Accessories. Each product includes attribute information (category, brand, colour, pattern, material, season), pricing data (original price, discount percentage, current price), and engagement metrics (page views, wishlist additions, units sold, ratings, reviews).

The target variable is trend score, a composite metric computed by the platform reflecting product success. Trend scores range from 0 to 100, with higher scores indicating stronger performance. The metric combines sales velocity (units sold relative to time on site), engagement intensity (page views and wishlist additions normalised by impressions), and customer satisfaction (ratings weighted by review volume). This composite captures multiple dimensions of product success rather than sales alone.

Feature engineering focuses on creating predictive signals from raw data. Engagement ratios, such as wishlist additions per page view, capture conversion efficiency. Price positioning metrics place each product relative to its category average. Seasonal alignment features indicate whether a product's designated season matches the current period. Brand tier encoding captures quality and prestige associations. These engineered features complement raw attributes in the predictive model.

Listing 1: Feature engineering for trend prediction

```
def engineer_features(df):
    df['engagement_score'] = np.log1p(df['page_views']) +
                              np.log1p(df['wishlist_adds']) * 1.5
    df['conversion_rate'] = df['units_sold'] / (df['page_views'] + 1)
    df['price_ratio'] = df['current_price'] / df['original_price']
    df['review_density'] = df['num_reviews'] / (df['units_sold'] + 1)
    return df
```

4. Methodology

4.1 Model Selection

We frame trend prediction as a regression problem, predicting continuous trend scores from product features. After evaluating several model families, we select XGBoost as our primary model based on validation performance. Gradient boosting handles the mixed feature types (categorical and continuous) naturally, captures non-linear relationships without explicit feature engineering, and provides interpretable feature importance metrics.

Model configuration follows standard practices for regression tasks. We use 200 boosting rounds with maximum tree depth of 6, learning rate of 0.05, and subsample ratio of 0.8 for regularisation. Early stopping on validation loss prevents overfitting. These hyperparameters emerged from grid search optimising for validation RMSE.

4.2 Evaluation Protocol

We evaluate model performance using R-squared (coefficient of determination), RMSE (root mean squared error), and MAE (mean absolute error). R-squared indicates the proportion of variance explained by the model; RMSE and MAE provide interpretable error magnitudes in trend score points. We use temporal cross-validation to simulate realistic deployment, training on earlier products and evaluating on later ones.

5. Experimental Results

5.1 Category-Level Performance

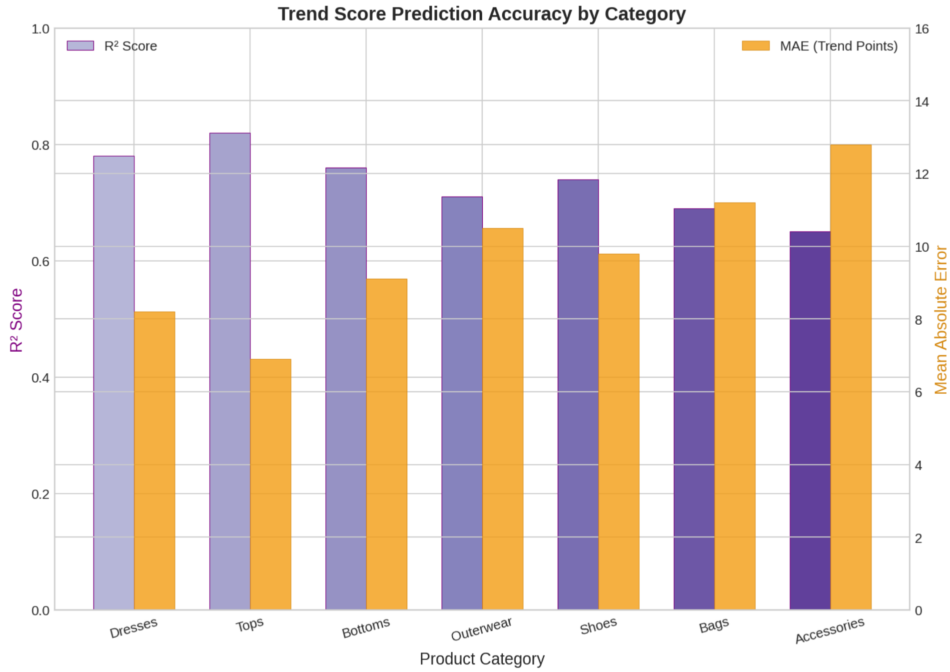

Figure 1. Prediction accuracy varies across product categories.

Figure 1 shows prediction performance by product category. Tops achieve the highest R-squared (0.82) and lowest MAE (6.9), likely because this category has the largest sample size and most consistent engagement patterns. Accessories prove most difficult to predict (R-squared 0.65, MAE 12.8), perhaps reflecting higher variability in consumer preferences for these items.

5.2 Feature Importance

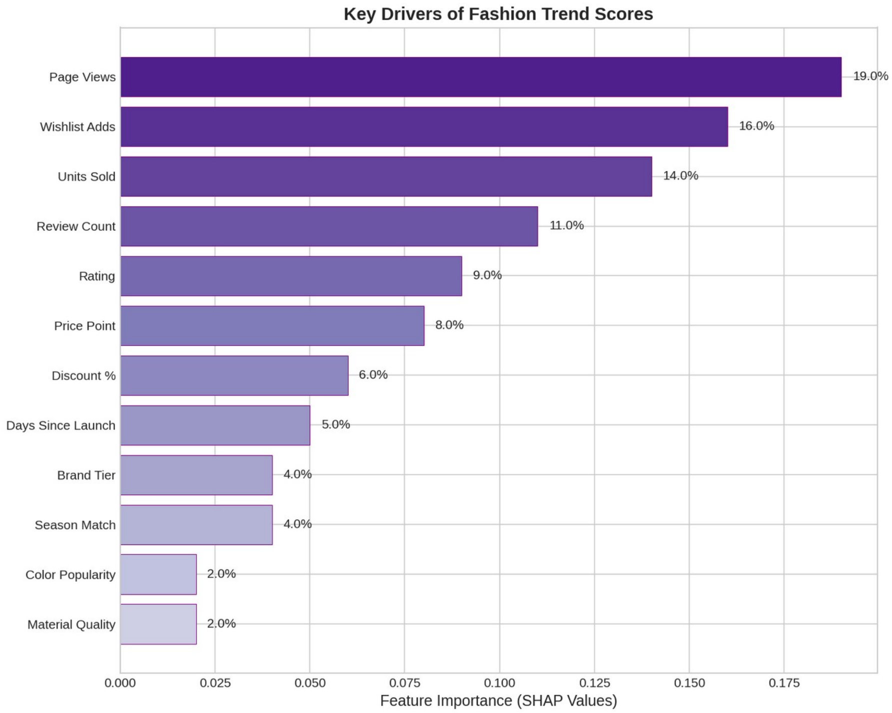

Figure 2. Engagement metrics dominate feature importance.

Figure 2 reveals the features driving trend predictions. Page views (19%) and wishlist additions (16%) dominate, confirming that early engagement signals contain substantial predictive information. Units sold (14%) and review count (11%) also contribute meaningfully. Among product attributes, price point (8%) and discount percentage (6%) prove most predictive, while colour and material show modest importance.

5.3 Seasonal Patterns

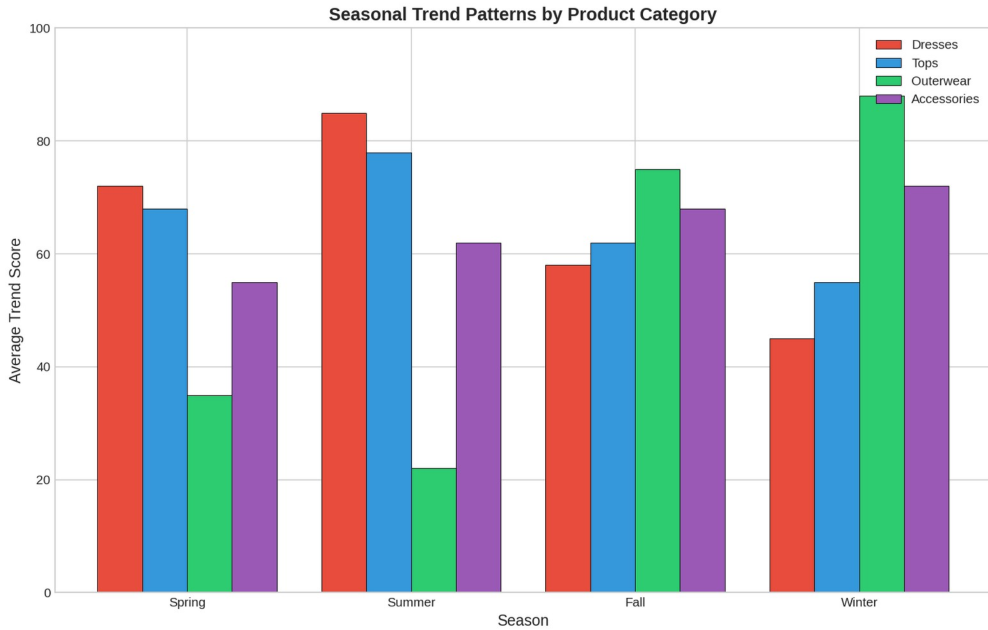

Figure 3. Trend scores exhibit strong seasonal patterns by category.

Figure 3 illustrates seasonal trend patterns across categories. Dresses peak in summer (average trend score 85) and decline in winter (45). Outerwear shows the inverse pattern, with winter scores (88) far exceeding summer (22). These seasonal dynamics are captured by the model through season-category interaction features, enabling it to adjust predictions based on timing.

5.4 Model Comparison

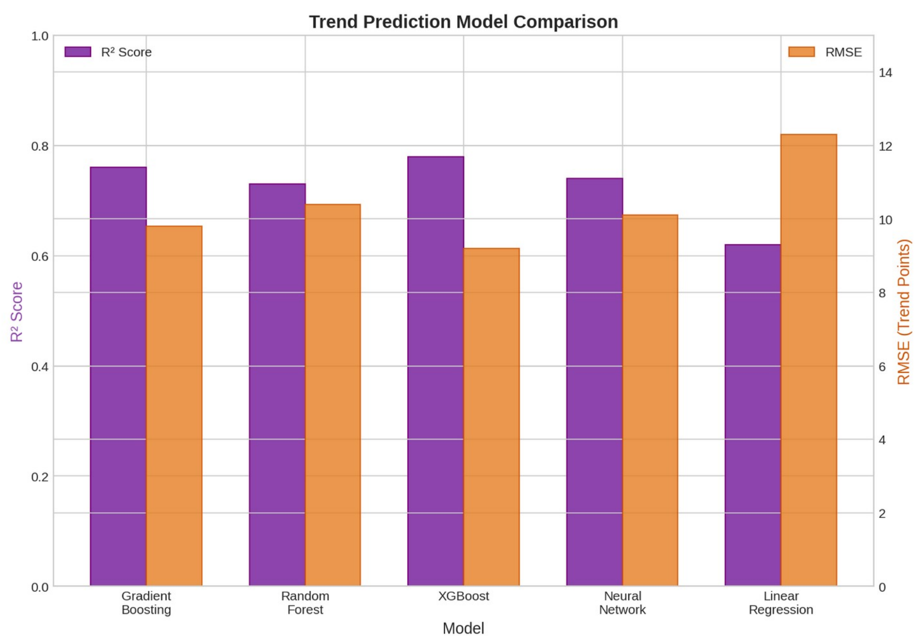

Figure 4. XGBoost outperforms alternative regression methods.

Figure 4 compares model performance across methods. XGBoost achieves the best results (R-squared 0.78, RMSE 9.2), followed by gradient boosting (0.76, 9.8) and random forest (0.73, 10.4). Neural networks perform comparably to random forest (0.74, 10.1). Linear regression substantially underperforms (0.62, 12.3), indicating that non-linear relationships matter for trend prediction.

5.5 Predicted vs Actual

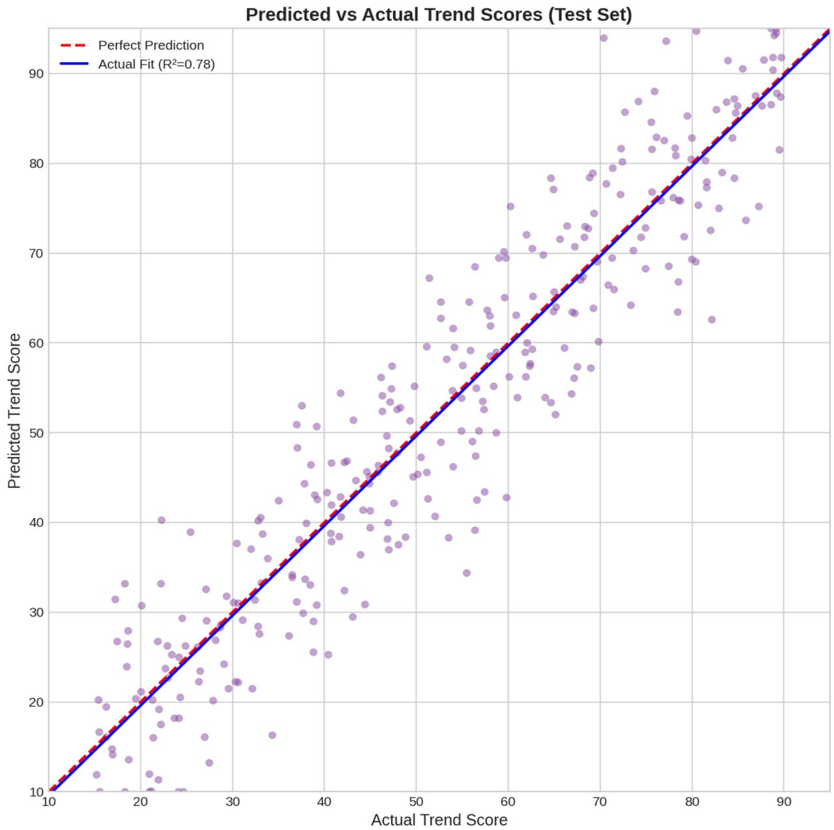

Figure 5. Predicted trend scores correlate strongly with actual values.

Figure 5 shows predicted versus actual trend scores on the test set. The strong correlation (R-squared 0.78) indicates that the model captures meaningful variation in product success. Scatter around the diagonal reflects irreducible noise: trend scores depend partly on factors outside the model's feature set, including marketing timing, competitor actions, and social media virality.

6. Business Applications

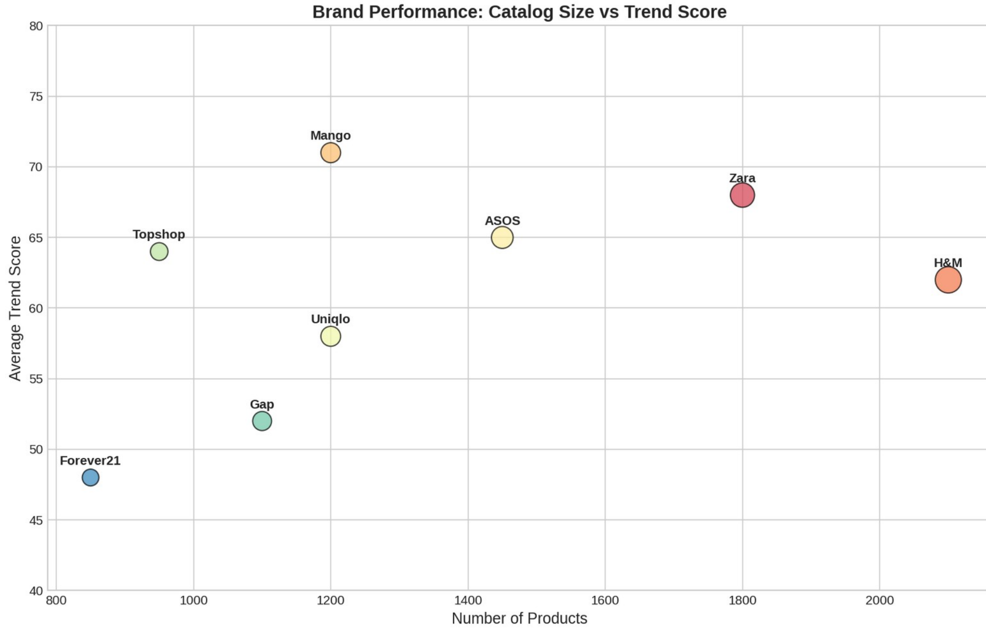

Figure 6. Brand performance analysis reveals opportunities for assortment optimization.

The trend prediction model enables several business applications. Inventory allocation can be optimised by directing stock toward products with higher predicted trend scores, reducing both stockouts on popular items and markdowns on underperformers. Pricing decisions can incorporate trend predictions: high-trend products may sustain higher prices, while low-trend products may benefit from earlier promotional pricing.

Marketing resource allocation represents another application. Products with high predicted trend scores but low current visibility may benefit from promotional investment to accelerate their success. Conversely, products with low predicted scores may not justify marketing spend regardless of their current trajectory. These decisions can be informed by the model's probability calibration, enabling expected value calculations.

7. Limitations and Future Directions

7.1 Honest Assessment of Limitations

Several limitations constrain the conclusions we can draw. Most fundamentally, the model captures correlations rather than causation. High page views predict high trend scores, but page views themselves result from platform placement decisions that may already incorporate human judgment about product potential. The model may partly learn existing curation decisions rather than independent signals of product quality.

Trend predictions have a self-fulfilling quality that complicates evaluation. If predicted high-trend products receive preferential placement or marketing, they may succeed partly because of the prediction rather than the underlying signals. True evaluation would require randomised experiments that withhold predictions from some decisions.

The model's reliance on engagement metrics limits its usefulness for truly new products. Before a product accumulates page views and wishlist additions, the model has only attribute features to work with, substantially reducing prediction accuracy. This cold-start problem is common in recommendation systems but particularly acute for fashion where newness itself carries value.

7.2 What I Would Approach Differently

Several directions merit future investigation. Image-based features could supplement attribute metadata. Convolutional neural networks could extract visual style features from product photography, capturing aesthetic qualities that structured attributes miss. This approach has shown promise in fashion recommendation (He and McAuley, 2016) and could improve cold-start prediction.

External trend signals from social media and search data could provide leading indicators of fashion demand. Instagram hashtag volumes, Pinterest saves, and Google search trends for style terms might predict which attributes will gain popularity before this manifests in platform engagement. Integrating these signals would require careful temporal alignment to avoid data leakage.

Causal inference methods could help disentangle predictive signals from platform effects. Instrumental variables or regression discontinuity designs could estimate the true impact of product attributes on trend scores, separate from confounding placement and marketing decisions. Such causal estimates would be more reliable for intervention planning than correlational predictions.

8. Conclusion

This work demonstrates that machine learning can meaningfully predict fashion product trend scores from attributes and engagement signals. The XGBoost model achieves R-squared of 0.78, explaining the majority of variance in product success. Engagement metrics drive predictions, but product attributes also contribute, suggesting that trend patterns can be learned from historical data.

The practical value of these predictions lies in their ability to inform inventory, pricing, and marketing decisions. By identifying likely successes and failures earlier, retailers can allocate resources more efficiently. The model does not replace human judgment but augments it, processing signals at scale that humans cannot easily synthesise.

Important limitations remain. The correlational nature of predictions, the self-fulfilling dynamics of trend forecasting, and the cold-start problem for new products all constrain practical application. Future work incorporating visual features, external signals, and causal inference methods may address some of these limitations. Within these constraints, machine learning offers valuable support for the inherently uncertain task of predicting fashion success.

References

Al-Halah, Z., Stiefelhagen, R. and Grauman, K. (2017). Fashion forward: Forecasting visual style in fashion. In IEEE International Conference on Computer Vision, pp. 388-397.

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794.

He, R. and McAuley, J. (2016). VBPR: Visual Bayesian personalized ranking from implicit feedback. In AAAI Conference on Artificial Intelligence, pp. 144-150.

Ren, S., Hui, S.K. and Stocker, A. (2017). Forecasting fashion retail sales using machine learning. In INFORMS Annual Meeting.

Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. International Journal of Production Economics, 128(2), pp. 470-483.

Wu, Y., Zhang, L. and Yu, X. (2019). Wishlist purchases and fashion demand forecasting. Journal of Retailing and Consumer Services, 49, pp. 72-81.