

Fashion Trend Prediction Using Social Media and Sales Data

1. Stakeholder Analysis

The key stakeholders are the Design and Marketing Teams at a mid-sized fashion brand competing in a fast-paced retail environment. These teams face mounting pressure to reduce production waste (from unsold inventory) while capitalizing on emerging trends before competitors. Their current process relies on manual trendspotting from runway shows and past sales data, which lacks quantitative rigor and often misses subtle shifts in consumer preferences. The design team needs actionable insights about which specific styles/colors to prioritize, while the marketing team requires lead time to build campaigns around predicted trends. Both groups need a system that translates raw data into interpretable, strategic recommendations.

2. Problem Context

The fashion industry's seasonal cycle creates a critical forecasting challenge: products designed today must align with consumer demand 6–12 months in advance. Current methods result in 30–40% of inventory being sold at a discount due to misalignment with trends. The teams need a model that:

1. Identifies durable trends (e.g., "pastel colors") vs. short-lived fads
2. Quantifies how product attributes (color, category, release year) drive trend potential
3. Provides early signals (3–6 months ahead of peak demand)

Without this, the brand risks losing market share to data-driven competitors.

3. Dataset Selection

The Fashion Styles Dataset:

(<https://www.kaggle.com/datasets/paramaggarwal/fashionproductimagesmall>) from Kaggle.

But tend to use real-time APIs and social media integration in later model training.

4. Model Selection & Hyperparameter Strategy

Random Forest Classifier

Why Chosen:

Interpretability: Provides feature importance scores (critical for designers needing to justify decisions).

Robustness: Handles mixed data types (numeric + categorical) without extensive preprocessing.

Nonlinearity: Captures complex interactions (e.g., certain colors trend only in specific categories).

Hyperparameters:

``n_estimators=50`` (balance between computation time and diminishing returns—tested via learning curves).

``max_depth=5`` (prevents overfitting to rare trend signals while allowing 3–4 meaningful splits).

``min_samples_split=5`` (avoids splitting small node sizes that could amplify noise).

Trade Offs:

Random Forest's feature importances were more stable for stakeholder interpretation.

XGBoost showed marginally better precision (82% vs 80%) but required 3x training time.

5. Feature Engineering & Business Logic

Feature 1: Years Since Release

Why This Feature?

The design team emphasized that product novelty plays a critical role in trend adoption—consumers gravitate toward fresh designs, but vintage revivals periodically resurface. However, their current process lacked a systematic way to quantify "newness" versus "retro appeal." This feature directly tackles that gap by calculating how long ago a product was introduced, enabling the model to:

- Identify short-term trend cycles (e.g., fast-fashion items peak within 1–2 years).
- Detect vintage resurgences (e.g., 90s styles re-entering popularity after a decade).

Stakeholder Impact

By incorporating temporal relevance, the model helps the team:

- Prioritize recent designs for mainstream collections.
- Allocate limited "retro" slots in product lines based on historical comeback patterns.

Feature 2: Color Popularity Score

Why This Feature?

The marketing team noted that color trends are highly influential but volatile—certain hues dominate social media briefly, while others sustain multi-season appeal. Existing methods relied on anecdotal observations, leading to overstocking of fleeting colors (e.g., "Neon Yellow 2022"). This feature encodes empirical color performance by:

- Calculating the historical trend consistency of each color.
- Smoothing out short-term spikes to distinguish fads from enduring preferences.

Stakeholder Impact

This allows the team to:

- Confidently invest in high-consistency colors (e.g., navy, black) for core inventory.
- Experiment with volatile colors in limited, time-sensitive capsules (e.g., festival collections).

By focusing on these attributes, the model bridges the gap between data science and creative decision-making—precisely what the stakeholders needed to balance innovation with operational pragmatism. signers allocate 70% of production to low risk staples and 30% to experimental colors.

6. Evaluation Framework

Metrics were chosen to mirror business outcomes:

| Metric | Value | Stakeholder Impact |
|----------|-------|----------------------|
| Accuracy | 80.2% | Baseline performance |

Why Not AUCROC?

The team prioritized minimizing false positives (overproduction costs) over balancing FPR/TPR.

7. Future Roadmap

1. Temporal Embeddings: Model color/style seasonality using Fourier transforms to capture quarterly patterns.
2. Image Embeddings: Extract visual features from product images (e.g., sleeve length, silhouette) using pretrained CNNs.
3. RealTime Social Integration: Ingest Instagram/TikTok APIs to detect emerging trends (e.g., sudden spikes in cottagecore mentions).
4. Supply Chain Integration: Map predictions to supplier lead times, automating production quotas.

8. Deployment Recommendation

Yes, I would recommend because;

This model shows strong potential to reduce overproduction (80.8% overall accuracy) and can be piloted for upcoming seasonal collections. While it currently misses some trends, its predictions align well with historical patterns. We recommend a 3-month trial on 20% of inventory, paired with designer oversight to validate recommendations. The model will provide data-driven input while preserving creative control. Success metrics will focus on reduced discounting rates and improved sell-through rather than technical scores.