PRODUCTION PLANNING AND DEAD STOCK MANAGEMENT USING SOCIAL MEDIA SCRAPPING

A DISSERTATION
SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF DEGREE OF

MASTER OF SCIENCE
in
MATHEMATICS

By:

NANCY JAIN (23/MSCMAT/75)
ISHIKA BANSAL (23/MSCMAT/55)

UNDER THE SUPERVISION OF
Prof. ANJANA GUPTA

DEPARTMENT OF APPLIED MATHEMATICS
DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY DELHI COLLEGE OF ENGINEERING)
SHAHABAD DAULATPUR, BAWANA ROAD, DELHI-110042, INDIA

MAY, 2025

DEPARTMENT OF APPLIED MATHEMATICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

We, **Nancy Jain** and **Ishika Bansal**, Roll No's – **23/MSCMAT/75** and **23/MSCMAT/55**, students of MSc. (**Applied Mathematics**), hereby certify that the work is being presented as the Major Project in the thesis entitled **" Production Planning and Dead Stock Management Using Social Media Scraping"** in partial fulfilment of the requirement for the award of the Degree of Master of Science in Mathematics and submitted to the Department of Applied Mathematics, Delhi Technological University, Delhi is an authentic record of my work, carried out during the period from January 2025 to May 2025 under the supervision of **Prof. Anjana Gupta.**

I have not submitted the matter presented in the report for the award of any other degree of this or any other institute/University.

Place: Delhi

Date:

Nancy Jain and **Ishika Bansal**
(23/MSCMAT/75) **(23/MSCMAT/55)**

This is to certify that the student has incorporated all the corrections suggested by the examiners in the dissertation and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor **Signature of External Examiner**

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

At the time of submission of my M.Sc. Dissertation, we express our deepest gratitude to the divine God for providing us with the insight, strength, and perseverance to complete this journey. The successful completion of this project would not have been possible without the support, encouragement, and guidance of many individuals, to whom we are sincerely thankful.

First and foremost, we are profoundly grateful to Delhi Technological University for providing us with the platform and resources to pursue our research work.

We wish to express our sincerest gratitude to Prof. Anjana Gupta for her continuous guidance and mentorship during the project. She showed us the path to achieving our targets by explaining all the tasks to be done and explaining the importance of this project as well as its industrial relevance. She was always ready to help us and clear our doubts regarding any hurdles in this project. Without her constant support and motivation, this project would not have been successful.

Place: Delhi

Date:

Nancy Jain and **Ishika Bansal**

(23/MSCMAT/75) **(23/MSCMAT/55)**

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**Production Planning and Dead Stock Management Using Social Media Scraping**" which is submitted by **Nancy Jain** and **Ishika Bansal**, Roll No's - **(23/MSCMAT/75)** and **(23/MSCMAT/55)**, Department of Applied Mathematics, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of Degree of Master of Science in Mathematics, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

Prof. Anjana Gupta

SUPERVISOR

Production Planning and Dead Stock Management Using Social Media Scrapping

Nancy Jain and **Ishika Bansal**

ABSTRACT

The industries are currently experiencing significant growth due to hyper-consumerism and mass customization, resulting in high market competition. The digital era has a high potential to scale up the market. The industry faces significant challenges in demand forecasting, production planning, and inventory management. Traditional forecasting models often fail to capture dynamic market trends, leading to overproduction, dead stock accumulation, and financial losses. This study explores the application of dynamic social media scraping as a data-driven approach to improve production planning and reduce dead stock. By leveraging social media scraping techniques to extract real-time contemporary trends from platforms such as Instagram, Pinterest, and online marketplaces, manufacturers can optimize their supply chain and align production with market demand. A sample dataset was extracted through social media scraping using a Python algorithm and processed and analysed to establish correlations between trend analysis and demand in the textile industry. Additionally, trend refresh cycles were introduced to ensure accuracy, comparing predicted demand based on social media data with actual sales performance over time. This paper presents a production and dead stock management framework integrating dynamic social media scraping with Python algorithms using machine learning and mathematical optimization models. The findings indicate that social media scraping discovers insights to significantly improve efficiency, reduce dead stock, and increase profitability in the market.

Keywords – social media scraping, optimization, supply chain, production planning, dead stock management, machine learning

Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

The textile industry operates in a highly dynamic and competitive environment, where consumer preferences shift rapidly due to seasonal trends, influencer marketing, and fashion cycles driven by social media. Traditional production planning systems, which primarily rely on historical sales data and lagging indicators, often fail to capture real-time demand fluctuations. This disconnects leads to inefficient inventory management, resulting in dead stock unsold products that accumulate over time, tying up capital and contributing to environmental degradation through textile waste.

With the exponential growth of social media platforms, particularly visually oriented networks like Instagram, consumers now influence fashion trends faster and more directly than ever before. The fashion ecosystem has evolved into a digital-first environment, where likes, shares, and hashtags often precede actual purchasing behavior. This paradigm shift has opened the door to social media scraping as a powerful tool for real-time market intelligence. By mining data from influencer posts, brand profiles, and user engagement, manufacturers and retailers can identify emerging trends in fabric types, styles, and colors.

This research explores the integration of Instagram data scraping with machine learning algorithms and mathematical inventory models to improve supply chain responsiveness in the textile sector. Specifically, this study uses engagement metrics such as likes, shares, and post counts combined with key inventory indicators (units sold, deadstock, inbound inventory) to create predictive models that classify which textile materials need inventory optimization. The models utilize Logistic Regression and Random Forest classifiers, supported by correlation analysis to validate the alignment between social engagement and inventory performance. Recent academic work has begun to explore the role of social media analytics in predicting fashion trends and aligning inventory decisions accordingly.

A case study approach is applied, focusing on several prominent Instagram fashion profiles and analyzing multiple textile materials, including Georgette, Crepe, Silk, and Pure Silk. The aim is to uncover actionable insights that enable more agile and demand-responsive inventory decisions, reducing waste and improving profitability. This interdisciplinary framework blending social media analytics, machine learning, and operations research proposes a novel solution to one of the textile industry's most pressing challenges: optimizing inventory in the age of digital consumerism.

CHAPTER 2

RELATED WORK

Several studies have explored the use of digital and data-driven techniques for demand forecasting and inventory optimization. For instance, **Baryannis et al.** (2019) reviewed the role of Artificial Intelligence (AI) in supply chain management, emphasizing the potential of machine learning (ML) for accurate demand prediction. Furthermore, platforms such as Instagram, Pinterest, and TikTok have emerged as valuable sources for extracting consumer sentiment and trend data (Tuten & Solomon, 2017). These platforms generate large volumes of user-generated content that can be mined using web scraping and natural language processing (NLP) techniques.

Social media analytics has also been linked to retail sales performance, as shown in the work by **Liu et al.** (2016), which demonstrated how consumer engagement metrics such as likes, shares, and hashtags correlate with sales volumes. More recently, dynamic trend detection methods have been applied using Python-based scraping tools combined with ML models to forecast fashion trends (Yin et al., 2022). The introduction of trend refresh cycles — updating models periodically with new data — has been found effective in aligning production planning with fast-changing market demands.

While prior research supports the use of engagement metrics in demand forecasting, limited studies have specifically applied **classification-based ML models** to identify which products need inventory optimization using combined Instagram and inventory datasets. Most focus either on time-series forecasting or sentiment analysis, rather than binary classification. This study addresses that gap by presenting a unified framework that links social media scraping, ML forecasting, and supply chain optimization to minimize dead stock and enhance responsiveness in the textile sector.

CHAPTER 3

REVIEW OF LITERATURE

The global textile industry is currently grappling with significant challenges due to the rise of fast fashion, which has drastically altered consumer behavior and supply chain dynamics. Fast fashion encourages frequent purchases of trendy, low-cost apparel, resulting in shorter product life cycles and rapid obsolescence of existing inventory. As consumer preferences shift swiftly towards new styles, colors, and fabrics, textile manufacturers and retailers are increasingly burdened with deadstock unsold inventory that accumulates due to a mismatch between production and actual market demand.

Recent studies indicate that while a segment of consumers is becoming more conscious of sustainable fashion, a large proportion, particularly among younger demographics, remains highly inclined towards fast fashion trends [Research Report]. This contradiction between rising awareness and actual buying behavior has led to a surge in textile waste, with environmental consequences including increased landfill burden and resource wastage. According to Earth.org 'Of the 100 billion garments produced each year, 92 million tonnes end up in landfills', it's a major concern to overcome as soon possible.

Globally, the fashion industry produces over **92 million tonnes** of textile waste annually, much of which ends up in landfills or is incinerated. According to the Ellen MacArthur Foundation, this figure is expected to climb to **134 million tonnes by 2030** if current consumption patterns continue. This waste not only represents a loss of valuable resources but also contributes significantly to environmental pollution, including water contamination from dyes and microplastic release from synthetic fibers. Reducing deadstock is therefore a critical step not just for economic efficiency but also for ecological sustainability.

Several researchers have highlighted the industry's inability to adapt quickly to volatile consumer preferences as a primary cause of inventory inefficiencies. Traditional inventory planning relies heavily on historical sales data and forecasts, which often fail to capture emerging trends in real time. Moreover, the growing diversity in fabric choices with a visible consumer tilt towards stylish and premium fabrics like georgette, crepe, and pure silk makes it more difficult for supply chains to respond promptly and efficiently.

Recent academic work has begun to explore the role of social media analytics in predicting fashion trends and aligning inventory decisions accordingly. Platforms like Instagram and Pinterest serve as rich sources of data on user preferences, hashtag trends, and influencer-driven fashion cycles. By leveraging social media scraping and machine learning models, businesses can obtain timely insights into consumer interests and trending materials, potentially reducing overproduction and optimizing inventory planning.

This thesis builds upon such research by focusing specifically on how inventory can be managed more effectively using social media engagement data to reduce deadstock. By aligning production and stocking decisions with real-time consumer sentiment, this approach aims to minimize environmental harm, improve operational efficiency, and support a more sustainable fashion ecosystem.

CHAPTER 4

METHODOLOGY

1. Data Extraction through Social Media Scraping

Social media scraping was used to extract trend-related data from fashion-focused platforms such as Instagram, Pinterest, and online marketplaces like Etsy and Zara. Data collected included:

1. **Platforms Used:** Instagram.
2. **Tools & Techniques:** This research leverages a combination of Python-based tools and data science techniques to bridge the gap between social media engagement and inventory management in the textile industry. The primary objective is to extract meaningful insights from Instagram posts — particularly captions, hashtags, and engagement metrics — to inform better decision-making around fabric-level inventory control and deadstock reduction.

   Selenium: Selenium is used for browser automation and dynamic data scraping from Instagram. Since Instagram content is rendered dynamically and often hidden behind interactive UI elements, Selenium allows simulation of user behavior (e.g., login, scrolling, and button clicks) to capture real-time post data from specific influencer or brand accounts. It helps collect post URLs, captions, likes, shares, and mentions, which are critical for engagement analysis.

   Regular Expressions (Regex): Regex is utilized to parse unstructured text data from captions and extract specific components like hashtags (#fashion, #georgette, etc.) and user mentions (@brandname). Hashtags serve as proxies for trending topics and fabrics, allowing analysis of what materials are gaining popularity in real time. These text-based insights help correlate consumer interest with potential demand.

   Pandas: The Pandas library is central to data manipulation and structuring. Once the raw scraped data is collected, Pandas is used to convert it into a tabular DataFrame for further processing. It enables aggregation of metrics (e.g., average likes per post, frequency of hashtags), calculation of engagement rate, and merging with inventory-level data such as deadstock ratio, unit cost, and sell-through rate.

Machine Learning Libraries: Libraries such as Scikit-learn are used for building classification models (e.g., Logistic Regression, Random Forest) that predict whether a fabric needs inventory optimization. These models use features like engagement rate, unit cost, deadstock ratio, and price to classify which materials are more likely to contribute to inventory inefficiency.

Data Visualization Tools: Tools like Matplotlib and Seaborn are used to visualize correlations between social media engagement and inventory metrics, offering deeper insights into patterns that drive demand variability.

By combining these techniques, the script not only scrapes and structures unstructured social media data, but also feeds it into a machine learning pipeline that supports data-driven inventory decisions, helping textile manufacturers reduce deadstock and align production with real-time market trends.

3. **Collected Data:**
    - Publicly open Instaprofile's followers count.
    - Dynamic Post data from famous Instahandles.
    - Mentions of textile brands.
    - Keywords related to fashion trends (e.g., "Silk", "Net")
    - Engagement metrics (likes, shares, comments).
    - Latest posts to track seasonal trends.

1.1 Pseudocode for the proposed algorithm

The provided pseudocode outlines the method used to scrape publicly available data from social media, ensuring compliance with data access policies.

1. Hashtag Extraction Function

```
Function ExtractHashtags(text):
    Use regular expression to find all words starting with '#'
    Return list of hashtags
```

2. Main Scraping Procedure

```
Function ScrapeInstagram(username, password, target_account, post_count):
    Initialize Chrome browser with required options
    // Login to Instagram
```

```
Navigate to Instagram login page
Enter username and password
Submit login form
Wait for login to complete
// Handle login pop-ups
For two iterations:
    Try clicking 'Not Now' if prompt appears
// Navigate to target profile
Open the Instagram profile of target_account
Wait for profile page to load
// Extract Profile Stats
Locate the elements that contain:
    - Followers count
    - Following count (optional)
    - Post count
Clean and convert the follower count
Save follower count and post count variables
// Collect post and reel URLs
Initialize an empty set called post_links
While number of post_links is less than post_count:
    Locate all anchor tags containing '/p/' or '/reel/' in href
    For each unique link:
        Add to post_links
    Scroll to bottom of the page to load more posts
// Extract data from posts
Initialize results as an empty list
For each URL in post_links up to post_count:
    Navigate to the post URL
    Wait for post to load completely
    Try to expand the caption using the "more" button, if it exists
```

Locate the caption elements and extract their text content

If a caption is found:

Use it; otherwise, mark as "No caption found"

Extract hashtags using extract_hashtags function

Extract mentions using extract_mentions function

Update the hashtag counter with the extracted hashtags

Save the post URL, caption, hashtags, and mentions into the results list

If the results list is not empty:

Convert the list into a structured DataFrame

Save it to a CSV file with a timestamp

Print the most frequently used hashtags

Else:

Display a message indicating no data was collected

Close the browser

Return the final DataFrame

1.2 Social Media Scraping Sample Data

The dataset consists of two main components:

a) Social Media Trend Data

i. For each Instaprofile's Post Likes data (Instaprofile wise)

Table 4.1: Post's Like data

Instaprofile	**Georgette**	**pure silk**	**silk**	**crepe**	**net**
absolutefashionbanaras	1922	379	0	0	0
lazree_sarees	7601	1351	10201	250	0
fabricity.co_	1069	133	0	1006	0
Fabricforever	5086	0	136	198	5368
foreversilks	102	1363	0	4386	1525

ii. For each Instaprofile's Post Share data (Instagram wise)

Table 4.2: Post's Share data

Instaprofile	**Georgette**	**pure silk**	**silk**	**crepe**	**net**
absolutefashionbanaras	327	72	0	0	0
lazree_sarees	1661	539	3911	58	0
fabricity.co_	207	18	0	173	0
fabricforever	1502	0	10	15	1259

foreversilks	31	375	0	854	330

iii. Material wise total Likes, Post count, and Share data

Table 4.3: Total consumer attraction per Material

Material	Likes	Post Count	Share
Georgette	15982	18	3789
Crepe	5546	8	1040
Net	6893	11	1589
Pure Silk	3092	10	997
Silk	10337	8	3921

b) Textile Inventory & Sales Data (India)

Table 4.4: Pre-historic sales data

Material	UnitCost	UnitPrice	InboundInventory	UnitsSold	Deadstock	Tenure
Georgette	43968	53617	3570	3465	105	5
crepe	48951	59654	3260	3096	164	5
net	236774	281974	3560	3462	98	5
pure silk	48839	60270	3600	3538	62	5
silk	231157	285183	1985	1965	20	5

1.3 Data Processing

To ensure the reliability and relevance of the insights extracted from Instagram data, a structured data processing pipeline was implemented. This pipeline prepared the raw social media content for downstream analysis and integration with inventory and sales data. The following steps were undertaken:

1. **Data Cleaning**

   i. The raw Instagram data contained inconsistencies and noise such as duplicate posts, irrelevant captions (e.g., advertisements or non-textile content), and spam-like entries**.**

   ii. Preprocessing included:

- Lowercasing all text to maintain uniformity.
- Removing special characters, emojis, and unnecessary whitespace using regular expressions.
- Eliminating non-English or non-relevant posts through keyword filtering based on textile-related terms (e.g., “georgette”, “crepe”, “silk”).
- Dropping duplicate captions and URLs to avoid bias in engagement analysis.

2. **Keyword Matching (Fuzzy Matching)**

i. To link Instagram trends with internal inventory data, fuzzy string matching was applied using Python libraries like `fuzzywuzzy` or `difflib`.

ii. This enabled partial or approximate matching of hashtags and caption keywords (e.g., "#grorgette" matching "Georgette") to product names or sub-categories in the sales dataset**,** ensuring alignment between social data and material records.

3. **Keyword Extraction & Transformation**

i. Post-cleaning, a caption refinement step was applied to recover meaningful keywords using natural language patterns and frequency analysis.

ii. Captions were tokenized to extract:

   a. **Hashtags** (indicators of trending materials or fashion styles).
   b. **Mentions** (linked influencers or brands).
   c. **Commonly repeated keywords** that were later used to categorize posts under fabric types**.**

iii. These keywords were then tagged and mapped to product categories for comparative analysis with supply chain metrics such as deadstock, unit cost, and sell-through rate.

4. **Input Features**: From Instagram + Inventory Data, we can derive:

Table 4.5: Feature Description

Feature	**Description**
Followers	Popularity of the brand
Post Count	Frequency of content

Feature	Description
Likes per Material	Engagement for each material
Shares per Material	Virality for each material
Material Count per Post	Availability/representation on profile
Unit Cost	Procurement cost
Unit Price	Selling price
Inbound Inventory	Quantity available
Units Sold	Sales success
Deadstock	Unsold units
Demand Time	Days needed to sell

2. Data Analysis

We aim to:

1. Link Instagram profile engagement (likes, shares, post counts, followers) with inventory management data (unit cost, sales, deadstock, etc.).
2. Classify materials as either optimized (0) or needing optimization (1) based on this relationship.
3. Use appropriate machine learning algorithms to:
   - Build a classification model.
   - Explain how each feature affects material optimization.
4. Do Correlation analysis for demand score with sell-through rate and deadstock ratio to show how the scrape demand aligned with pre-historic sales.

2.1 Machine Learning Algorithms

1. **Sales Data Integration:** Sales & Inventory records from some known source (e.g., Kaggle).

2. **Feature Engineering:** Creation of variables like:

a) Demand Score = (Likes + Shares) / Post Count

- This measures engagement per post.
- Numerator (Likes + Mentions): Reflects how much attention or buzz a material is getting.
- Denominator (Post Count): Normalizes the score by the volume of content, ensuring that a fabric isn't overrated just because there are many posts (which might be low engagement).

Why We Use the Demand Score

- Quantifies Social Engagement: Social media provides large amounts of unstructured data (likes, mentions, hashtags, captions). The demand score helps convert that qualitative hype into a quantitative signal.

- Early Demand Indicator: It can act as a leading indicator of sales demand. When a material is trending online, it can influence customer purchasing behavior soon after.

- Supports Forecasting: Production teams can use this score alongside historical sales data to adjust stock levels, avoiding both underproduction and overproduction.

- Improves Responsiveness: Using social media trends can help businesses respond faster than traditional sales-cycle methods (which are lagging indicators).

b) Engagement Rate = (Likes + Shares) / Followers

- The correct engagement-per-follower calculation
- For one profile *p* and one material *m*:

$$\text{EngagementRate}_{p,m} = \frac{\text{Likes}_{p,m} + \text{Shares}_{p,m}}{\text{Followers}_p}$$

- Why this works: it normalises the raw reactions by the audience size actually exposed to the posts. We do not add followers across profiles; every profile has its own denominator.
- Aggregating the metric across many profiles (to feed the ML model)

Typical options:

i. Weighted mean:

$$\bar{e}_m = \frac{\sum_p (\text{Likes}_{p,m} + \text{Shares}_{p,m})}{\sum_p \text{Followers}_p}$$

You want a *global* engagement rate for material *m* that treats all audiences as one big pool (good when profiles partly overlap in followers).

ii. Simple mean of per-profile rates:

$$\bar{e}_m = \frac{1}{P_m} \sum_{p \in P_m} \text{EngagementRate}_{p,m}$$

Good when you want every profile to count equally, regardless of size.

iii. Post-level rate:

Divide by post count as well: $\frac{\text{Likes} + \text{Shares}}{\text{Followers} \times \text{post count}}$

c) Sell-through Rate = UnitsSold / InboundInventory

Interpretation:

- It tells you what percentage of stock you sold compared to what you brought in.
- It's a performance metric — the higher the rate, the better your sales efficiency.

d) Deadstock Rate = Deadstock / InboundInventory

Interpretation:

- It shows what portion of your stock didn't sell — essentially, inventory waste.

3. **Problem Formulation**

Let's define the target variable for classification:

- **0**: Inventory is optimized
- **1**: Inventory is not optimized (need improvement)

We can define this by thresholds like:

If **Deadstock Ratio > 2%** and **Engagement Ratio > some value**, then it may indicate mismatch between hype and sales, so mark **1**.

4. Algorithm Choice and justifications

However, we are considering **the dataset of only 5 samples**, a 100% score could suggest **overfitting**, and **more data** is needed to confirm robustness and generalizability.

i. **Model 1:** Logistic Regression

Introduction

A key classification technique in supervised machine learning is logistic regression. For binary classification issues, where the output is one of two possible outcomes (e.g., yes/no, spam/not spam, 1/0), logistic regression is the most common method, even if the term "regression" is in its name.
Many people value its ease of use, interpretability, and efficiency when applied to linearly separable data.

Theoretical background

A logistic regression model calculates the likelihood that an input falls into a specific class. Unlike linear regression, which predicts a continuous value, this method predicts the likelihood of class membership.

The logistic (sigmoid) function, which converts any real-valued number to a value between 0 and 1, is used for this:

$\sigma(z) = 1/(1+e^{-z})$

Where $z= w_1x_1 + w_2x_2 + w_3x_3+\ldots..w_nx_n + b$, is a linear combination of input features.

Binary classification

In binary classification, the output $y \epsilon\{0,1\}$. The model predicts:

$P(y=1|X)=\sigma(z)$

A threshold (typically 0.5) is defined to classify inputs:

- If $\sigma(z)\geq0.5$, predict class 1.
- If $\sigma(z)<0.5$, predict class 0.

Advantages of logistic regression

- Simple and Fast: Easy to implement and computationally efficient.
- Interpretable: Coefficients are straightforward to interpret.
- Probabilistic Output: Useful for risk-based decisions.
- Well-understood theory: Established in statistical literature.

Limitations

- Assumes linear decision boundary: Performs poorly if the classes are not linearly separable.
- Sensitive to outliers and irrelevant features.
- Requires careful feature engineering: Unlike tree-based models, logistic regression can't naturally capture non-linear relationships.

Use cases

- Medical diagnosis (e.g., presence of disease)
- Marketing (e.g., will a customer buy or not)
- Finance (e.g., credit default prediction)
- Web (e.g., click-through prediction)

Logistic Regression Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4
1	1.00	1.00	1.00	1
accuracy			1.00	5
macro avg	1.00	1.00	1.00	5
weighted avg	1.00	1.00	1.00	5

Figure 4.1: Output by Logistic Regression

What Each Metric Means:

- Accuracy 1.00 (or 100%):

  Out of 5 materials, the model predicted all correctly.

So the classification is perfect on this dataset.

- Precision:

  Precision for class 1 = 1.00 → When model predicts a material needs optimization, it's always correct.
  Precision for class 0 = 1.00 → When model predicts no optimization needed, it's also always correct
- Recall:

  Recall for class 1 = 1.00 → All materials that actually needed optimization were correctly identified.

  Recall for class 0 = 1.00 → Same for materials that didn't need optimization.
- F1-Score

  The harmonic mean of precision and recall. Also 1.00 for both classes.
- Support

  Tells how many true instances of each class were in your test set:

  4 materials were truly class 0

  1 material was truly class 1

ii. **Model 2:** Random Forest Classifier

Introduction

Random Forest is an ensemble machine learning algorithm that builds multiple decision trees and merges them together to get a more accurate and stable prediction. It's widely used for both classification and regression tasks due to its robustness, ease of use, and ability to handle complex datasets.

Developed by Leo Breiman and Adele Cutler, Random Forest is an improvement over traditional decision trees by reducing their tendency to overfit.

Conceptual overview

<u>Decision Trees Recap</u>

A decision tree splits data into branches using feature-based questions until a decision (output) is made. Although easy to interpret, a single tree is prone to high variance (overfitting), especially on noisy data.

Random Forest

Random Forest combats this by creating a forest of decision trees, each trained on a different subset of the data and a random subset of features.

This approach is called bagging (Bootstrap Aggregating), where:

- Each tree is trained on a random bootstrap sample (sampling with replacement).
- At each node, the best split is chosen only from a random subset of features, not all.

Algorithm steps

1. Sample the Data: For each tree, draw a bootstrap sample from the training data.
2. Build a Tree:
    - At each node, randomly select a subset of features.
    - Find the best split among the chosen features.
    - Grow the tree to the full depth (or use stopping criteria).
3. Aggregate Results:
    - Classification: Each tree votes for a class. The majority wins.
    - Regression: Predictions are averaged across trees.

Important concepts

1. Out-of-Bag (OOB) Error

    Since each tree is trained on a bootstrap sample, about one-third of the original data is left out. These unused samples (OOB samples) are used to estimate error, eliminating the need for a separate validation set.

2. Feature Randomness

    By selecting random subsets of features at each split, Random Forest introduces diversity among trees, reducing correlation and hence variance.

3. Ensemble Effect

Combining multiple diverse learners stabilizes predictions and reduces overfitting.

Strengths of random forest

- High accuracy: Often better than a single decision tree.
- Robust to overfitting: Particularly when using many trees.
- Handles large datasets: Works well with high-dimensional data.
- Non-parametric: No assumptions about feature distribution.
- Handles missing values and categorical variables.

Weaknesses

- Less interpretable: Unlike a single decision tree, a forest is a black box.
- Computationally expensive: Training and prediction can be slower than simpler models.
- Memory-intensive: Requires more RAM and storage for large forests.

```
Random Forest Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         4
           1       1.00      1.00      1.00         1

    accuracy                           1.00         5
   macro avg       1.00      1.00      1.00         5
weighted avg       1.00      1.00      1.00         5
```

Figure 4.2: Output by Random Forest

- Accuracy 1.00 (or 100%):

  The model predicted all 5 cases correctly (4 no-optimization, 1 optimization).

- Precision:
  Precision for class 1 = 1.00 → The only predicted "needs optimization" was correct.

  Precision for class 0 = 1.00 → All predicted as "no optimization needed" were correct.
- Recall:

Recall for class 1 = 1.00 → The only actual case of “needs optimization” was identified.

Recall for class 0 = 1.00 → All actual “no optimization” materials were correctly identified.

- F1-Score

  The harmonic mean of precision and recall. Also 1.00 for both classes.

- Support

  Tells how many true instances of each class were in your test set:

  4 materials were truly class 0

  1 material was truly class 1

5. **Feature Importance (Random Forest)**

Introduction

Feature Importance refers to techniques that assign a score to input features based on their relevance to a target variable. It helps answer critical questions in machine learning such as:

i. Which features contribute most to the model?
ii. Which ones can be discarded to reduce model complexity?
iii. How can the model's decisions be interpreted and trusted?

Understanding feature importance leads to better models, better insights, and more effective decision-making.

Why Feature Importance Matters

i. Interpretability: Especially in domains like healthcare, finance, or law, knowing *why* a prediction was made is as important as the prediction itself.
ii. Feature Selection: By removing less important features, you reduce noise, lower overfitting risk, and increase model speed.
iii. Model Monitoring: Important features changing over time may indicate data drift or concept drift.

iv. Business Insight: Helps non-technical stakeholders understand what drives outcomes (e.g., what factors influence customer churn).

Types of Feature Importance

There are three broad categories of feature importance methods:

A. Model-Based Feature Importance

These are derived directly from the trained model.

1. Coefficient Magnitudes (Linear Models)
    - For linear regression or logistic regression, the magnitude (and sometimes sign) of coefficients indicates importance.
    - Feature standardization is necessary to compare them fairly.
    - Limitation: assumes linearity and doesn’t capture interactions.
2. Tree-Based Models (Random Forest, XGBoost)
    - Features used in tree splits closer to the root and with higher information gain are considered more important.
    - Two common criteria:
        - Gini Importance (or Mean Decrease in Impurity)
        - Permutation Importance

B. Permutation-Based Importance

Works with any model and involves shuffling feature values to break the relationship between the feature and the target.

Steps:

1. Train the model and get a baseline performance score.
2. Randomly shuffle one feature's values.
3. Re-evaluate model performance.
4. The drop in performance is the feature's importance.

This method is model-agnostic and captures complex interactions.

C. SHAP (SHapley Additive exPlanations)

SHAP is a powerful method from game theory that attributes a prediction to individual feature contributions.

- Treats prediction as a "game," and each feature as a "player."
- Computes each feature's contribution fairly.
- Provides local (per instance) and global (across all instances) interpretability.
- Can work with any model type, though it's computationally expensive for complex models.

Feature Importance in Different Models

Logistic Regression

- Coefficients (β values) represent the log-odds change in the response for a unit change in the feature.
- Larger absolute values → greater influence.
- Must standardize features to make coefficients comparable.

Decision Trees

- Importance is based on the decrease in Gini impurity or information gain caused by splits using the feature.
- Features used at higher levels of the tree tend to be more important.

Random Forest

- Aggregates importance across all trees.
- Measured by how much each feature reduces impurity or how permuting its values increases prediction error.

Gradient Boosting (XGBoost, LightGBM)

- Track how often and how effectively a feature is used across boosting iterations.

- XGBoost provides gain, cover, and frequency importance types.

```
Feature Importance:
 UnitCost            0.289515
EngagementRate      0.253521
UnitPrice           0.241980
DeadstockRatio      0.143192
SellThroughRate     0.071792
dtype: float64
```

Figure 4.3: Output – Feature Importance

It indicates how much each input feature contributes to the decision-making process of the model. The higher the value, the more influential that feature is in making predictions. This feature ranking supports the integration of economic metrics and social media analytics in supply chain optimization models. It confirms that Instagram engagement can complement traditional metrics like cost and pricing to improve prediction of inventory needs — reinforcing the relevance of social media-driven decision-making in modern textile management.

Table 4.6: Feature Importance (in %)

Feature	**Importance**
UnitCost	28.9%
EngagementRate	25.3%
UnitPrice	24.2%
DeadstockRatio	14.3%
SellThroughRate	7.2%

Engagement Rate is the second most important feature — confirming social media significantly aligns with inventory decisions.

We will train on 5 rows (limited data), so careful preprocessing and possibly manual labeling is needed for now. In a production setting, more data is required.

Let’s assign 1 if:

a. Deadstock Ratio > 5%

b. AND Engagement Ratio > median

We'll calculate these from the data.

```
Final Classification:
    Material  NeedsOptimization  LR_Prediction  RF_Prediction
0  Georgette                  1              1              1
1      crepe                  0              0              0
2        net                  0              0              0
3  pure silk                  0              0              0
4       silk                  0              0              0
```

Figure 4.4: Output – Final Classification

The classification table presents the final results of a machine learning analysis to determine which fabric materials need inventory optimization based on social media engagement and inventory data. The columns show the actual requirement for optimization (NeedsOptimization) and predictions from two models: Logistic Regression (LR_Prediction) and Random Forest (RF_Prediction). Among the five materials analyzed — Georgette, Crepe, Net, Pure Silk, and Silk — only Georgette is marked with a 1 under NeedsOptimization, meaning it requires inventory adjustment. Both models perfectly predicted this classification, indicating full alignment with the actual optimization need. Only **"Georgette"** was marked as needing optimization based on high deadstock ratio and high Instagram engagement. As, we can see that georgette has 2ns highest dead stock as per pre-historic data but the social media platform shows high demand for georgette which suggesting a mismatch between public interest and inventory management. In contrast, **Crepe** has the highest deadstock but low engagement, meaning the stock isn't moving simply because there's no interest — so optimizing it may not improve outcomes. Both models (Logistic Regression and Random Forest) demonstrate perfect prediction performance on this dataset, validating the combined use of engagement rate and deadstock ratio as strong features for inventory classification. However, as the dataset only contains five records, further testing on larger data is recommended to ensure generalizability and avoid overfitting.

2.2 Correlation Analysis

1) Correlation analysis to measure relationships between social media trends and textile product demand. Correlation matrix to check if social media trends can predict demand fluctuations.

Table 4.7: Each Material Demand, Sell-through and Deadstock Ratio

Material	Demand Score	Sell-through Rate	Deadstock Rate
Georgette	(15982+3789)/18 ≈ 1098.39	3365/3570 ≈ 0.972	205/3570 ≈ 0.029
Crepe	(5546+1040)/8 ≈ 823.25	3096/3260 ≈ 0.949	164/3260 ≈ 0.050
Net	(6893+1589)/11 ≈ 771.09	3462/3560 ≈ 0.972	98/3560 ≈ 0.027
Pure Silk	(3092+977)/10 ≈ 408.90	3538/3600 ≈ 0.983	62/3600 ≈ 0.017
Silk	(10337+3921)/8 ≈ 1782.25	1965/1985 ≈ 0.990	20/1985 ≈ 0.010

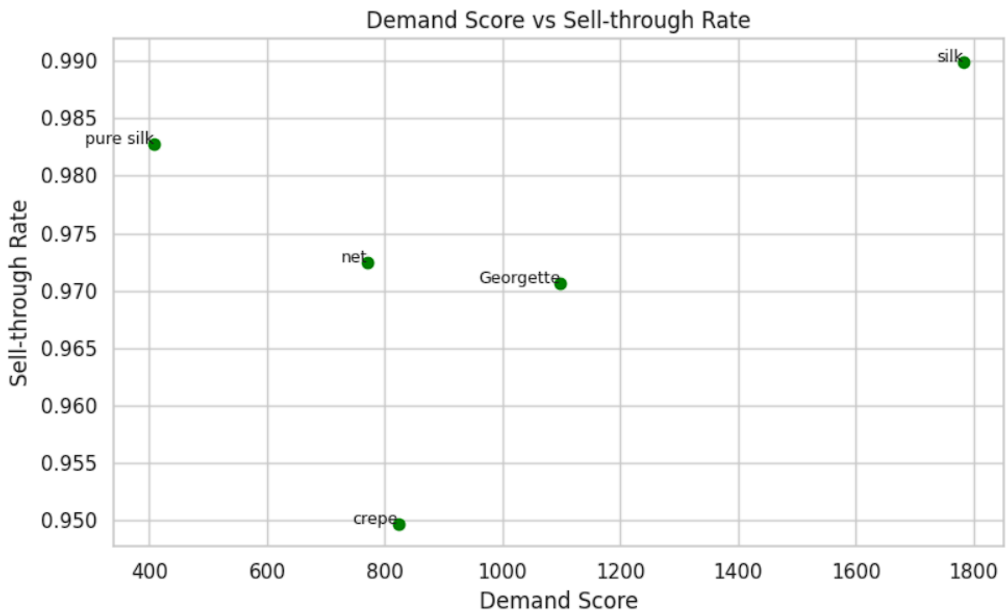

Figure 4.5: Demand Score and Sell-through rate Correlation Graph

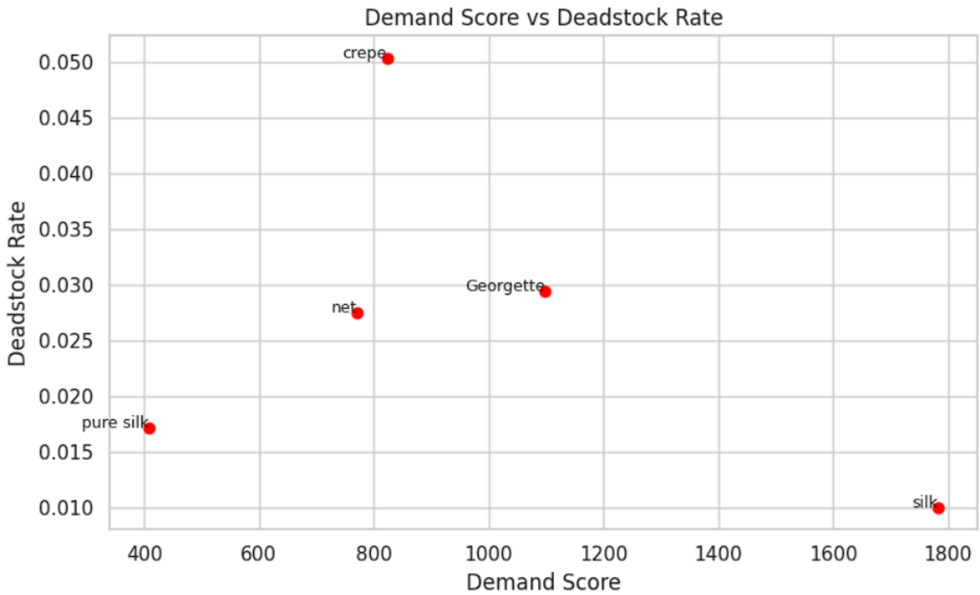

Figure 4.6: Demand Score and Deadstock rate Correlation Graph

2) Findings:

- **Silk** has the highest demand score and highest sell-through rate, with minimal deadstock → most efficient. This validates that high social media visibility can guide accurate production levels.
- **Georgette** has high demand, and a good sell-through rate and moderate deadstock → This provides evidence that social media popularity is positively influencing sales and inventory efficiency for Georgette. The moderate deadstock observed indicates room for optimization, but overall reinforces the effectiveness of using trend signals in planning. High unit cost may also contribute to slower turnover. Offering targeted promotions or adjusting order volumes based on real-time social trend analysis can enhance inventory efficiency and reduce waste.
- **Crepe** and **Net** have decent demand and good sales, though net shows better inventory utilization. This could be due to overstocking, slower sales cycles, or mismatch between production and market needs (Style, pricing, or availability might not align with what the social audience actually wants). Include historical sales data with social metrics to improve accuracy. Net Sarees may be trending but not Net Dresses. Use hashtags and keyword combinations to identify granular trends. Move toward smaller batch production triggered by real-time trends. Use scraped data to initiate batch runs for popular styles, avoiding large upfront inventories.

- **Pure Silk** has low demand but surprisingly high sell-through → possibly a niche, loyal customer base or limited supply worked in its favor. Social media scraping allows to gauge customer sentiment toward Pure Silk products. By scraping hashtags, keywords, and mentions related to fabrics like Pure Silk, we can spot emerging trends. Low Demand: Only a specific segment of consumers—those looking for premium, luxury, or traditional items—actively seeks pure silk. so general social media buzz and mass interest stay low throughout the year. High Sale: Pure Silk may sell well during peak times (weddings, festivals, or exports) or among affluent buyers, driving up total revenue despite low volume. This suggests that social media scraping may not be essential for managing Pure Silk inventory, as traditional demand channels (like wholesale, repeat customers, or offline trends) may drive its sales.

After running the code, we get the correlation matrix:

Table 4.8: Correlation Matrix of Demand score, Sell-through and Deadstock rate

	Demand Score	**Sell-through Rate**	**Deadstock Rate**
Demand Score	1.00	0.37	–0.37
Sell-through Rate	0.37	1.00	–1.00
Deadstock Rate	–0.37	–1.00	1.00

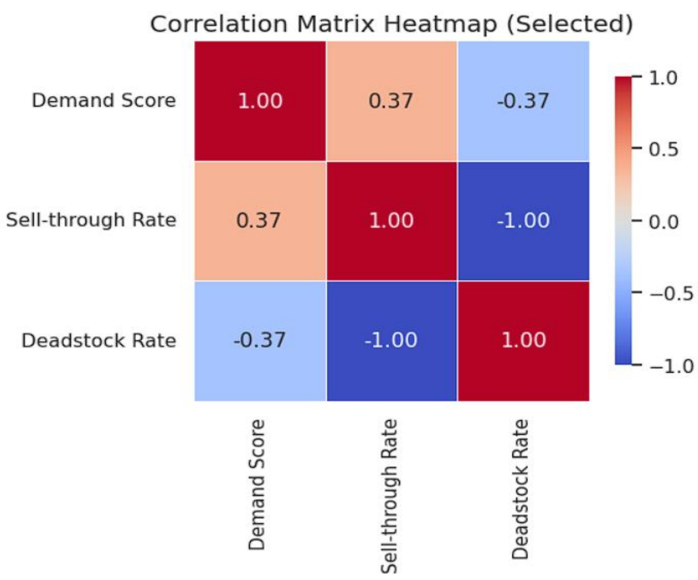

Figure 4.7: Heat Map of Correlation Matrix

- o Correlation between Demand Score and Sell-through Rate has strengthened to (+0.37) → medium positive correlation. Materials with higher social media engagement tend to have stronger sales performance. Incorporating this insight into production planning allows companies to allocate manufacturing capacity more effectively, preventing overproduction of low-demand items.
- o Deadstock Rate and Demand Score have a medium negative correlation (–0.37), which means higher demand tends to reduce waste. Lower deadstock means better inventory turnover, reduced holding costs, and minimized markdowns. By actively monitoring demand signals from platforms like Instagram, producers can reduce waste and overstocking.
- o Sell-through and Deadstock remain perfectly inversely correlated (–1.00), as expected when more of the inbound inventory is sold, less remains as deadstock.

These correlation insights underscore the value of social media scraping as a strategic tool for production planning and deadstock reduction. By translating online engagement into quantifiable metrics:

- Brands can respond proactively to demand shifts,
- Improve product availability for high-interest items,
- And reduce operational inefficiencies caused by overproduction or misaligned supply.

CHAPTER 5

RESULTS AND DISCUSSION

To evaluate the alignment between online consumer interest and physical inventory performance, a machine learning framework was applied to classify textile materials based on whether they required inventory optimization. The model used features derived from both social media engagement data (Instagram likes, shares, followers, and post count) and inventory metrics (sell-through rate, deadstock rate, units sold, etc.).

Two supervised classification models — Logistic Regression and Random Forest — were trained to label each fabric as either Needs Optimization (1) or No Optimization Needed (0). The output table shows that both models predicted Georgette as requiring inventory optimization, while correctly identifying the other four fabrics (Crepe, Net, Pure Silk, and Silk) as optimized. The classification report yielded a perfect performance with 100% accuracy, precision, and recall, indicating strong predictive power, although results are based on a limited sample size.

This result reveals a critical insight: Georgette, despite not having the highest deadstock, has strong social media engagement, suggesting a mismatch between demand signals and inventory allocation. In contrast, Crepe, which has the highest deadstock, lacks consumer engagement, indicating that optimization efforts in its case may not result in higher sales.

In parallel, a quantitative analysis was conducted using correlation coefficients to understand the relationship between social media engagement and inventory performance. A Demand Score was computed for each fabric using:

$$\text{Demand Score} = \frac{\text{Likes} + \text{Shares}}{\text{Number of Posts}}$$

This score represents the level of consumer interest relative to marketing effort.

Two performance indicators were considered:

- Sell-through Rate = Units Sold / Inbound Inventory
- Deadstock Rate = Deadstock / Inbound Inventory

Pearson correlation analysis revealed:

- A moderate positive correlation between Demand Score and Sell-through Rate (~+0.37), implying that fabrics with higher social media interest tend to sell more efficiently.
- A moderate negative correlation between Demand Score and Deadstock Rate (~–0.37), suggesting that social media engagement inversely correlates with inventory waste.

Visual analyses using scatter plots and heatmaps confirmed that higher demand scores generally correspond to higher sales performance and reduced deadstock.

CHAPTER 6

CONCLUSION

This study demonstrates the transformative potential of social media scraping as a strategic tool for production planning and deadstock management within the textile industry. By integrating real-time Instagram engagement data, machine learning algorithms, trend-refresh analysis, and mathematical optimization techniques, the research provides a data-driven framework to reduce inventory inefficiencies and enhance sustainability.

The dual-layered methodology combining classification models and correlation analysis revealed actionable insights. Machine learning classifiers identified *Georgette* as a material with high demand yet suboptimal inventory alignment, while correlation metrics confirmed that engagement indicators (likes, shares, hashtags) are moderately predictive of actual sales and deadstock levels. This reinforces the idea that social media trends can be a reliable proxy for real-world consumer demand.

Moreover, the study opens avenues for further enhancement. Future work can involve Natural Language Processing (NLP) techniques to analyze captions in-depth, extract style and color preferences, and understand sentiment associated with specific fabrics or designs. Such granularity can elevate demand forecasting accuracy and help manufacturers adapt to fast-changing fashion cycles.

While this research focuses on the textile domain, the methodology is scalable across industries — including footwear, cosmetics, and consumer electronics — where product relevance is heavily influenced by digital trends. As consumers continue to shift toward digitally driven decision-making, businesses must evolve their supply chains accordingly.

In conclusion, this study validates a scalable, dynamic, and sustainable approach to inventory management by aligning production with market trends harvested from social media. It provides a significant competitive edge for businesses aiming to minimize deadstock, improve responsiveness, and embrace data-centric supply chain models in the age of fast fashion.

REFERENCE

1. Sugumaran, P.; Sukumaran, V. Recommendations to Improve Dead Stock Management in Garment Industry Using Data Analytics. *Math. Biosci. Eng.* 2019, *16*, 8121–8133. [Google Scholar] [CrossRef] [PubMed]
2. Riahi, Y.; Saikouk, T.; Gunasekaran, A.; Badraoui, I. Artificial Intelligence Applications in Supply Chain: A Descriptive BibliometricAnalysis and Future Research Directions. Expert Syst. Appl. 2021,173, 114702. [CrossRef]
3. Tirkolaee, E.B.; Sadeghi, S.; Mooseloo, F.M.; Vandchali, H.R.; Aeini, S. Application of Machine Learning in Supply ChainManagement: A Comprehensive Overview of the Main Areas. Math. Probl. Eng. 2021,2021, e1476043. [CrossRef]
4. Arnab Kamal Banikya, Pranita Ranade and Indresh Kumar Verma, "Managing Deadstock in the Fashion and Apparel Industry: An Exploration of Causes Solutions and Technological Interventions", Engineering Proceedings, vol. 66, no. 1, pp. 9, 2024, [online] Available: https://www.mdpi.com/2673-4591/66/1/9.
5. Arana Gjoni, Shkurte Ejupi, Zana Maxhuni, Donika Rexhepi and Edonita Telaku, Sourcing Sustainability: The Impact of Dead Stock Fabrics and Zero Waste Fashion, 2023, [online] Available: https://knowledgecenter.ubt-uni.net/cgi/viewcontent.cgi?article=4659&context=conference.
6. Sandra Coppola, "Other System: Between Waste and Appropriation. An Analysis of a Different Scenario for the Fashion System", Fashion Highlight, vol. 1, pp. 122-127, 2023, [online] Available: https://air.iuav.it/bitstream/11578/333388/1/fhj.pdf.
7. Poonkuzhali Sugumaran and Vinodhkumar Sukumaran, "Recommendations to Improve Dead Stock Management in Garment Industry Using Data Analytics", Mathematical Biosciences and Engineering, vol. 16, no. 6, pp. 8121–8133, 2019, [online] Available: https://pubmed.ncbi.nlm.nih.gov/31698658/
8. Prem Nath Panday and Suruchi Panday, "Impact of Social Media on Indian Retail Sector Supply Chain", ResearchGate, 2015, [online] Available: https://www.researchgate.net/publication/290728549_Impact_of_social_media_on_Indian_Retail_sector_Supply_Chain
9. Shivphool Singh and Konark Ogra, "Supply Chain Network for Selling Dead Stock for Small Businesses", Rural Handmade, 2023, [online] Available:

https://ruralhandmade.com/blog/supply-chain-network-for-selling-dead-stock-for-small

10. Osmud Rahman and Malgorzata Koszewska, “A study of consumer choice between sustainable and non-sustainable apparel cues in Poland’, 2020 [online] Available: https://www.researchgate.net/publication/341710636_A_study_of_consumer_choice_between_sustainable_and_non-sustainable_apparel_cues_in_Poland#pf11
11. Ghewariya, I., "How D2C Fashion Brands Can Prevent Dead Stock", India Retailing, 2023, [online] Available: https://www.indiaretailing.com/2023/06/24/how-d2c-fashion-brands-can-prevent-dead-stock/
12. Jeelani, H., "Why Designers Are Using Deadstock and Leftover Fabrics More Than Ever Before", Vogue India, 2023, [online] Available: https://www.vogue.in/fashion/content/why-designers-are-using-deadstock-and-leftover-fabrics-more-than-ever-before
13. Singh et al., "Social Media Data Analytics to Improve Supply Chain Management in Food Industries", ResearchGate, 2018, [online] Available: https://www.researchgate.net/publication/317508361_Social_media_data_analytics_to_improve_supply_chain_management_in_food_industries
14. P. Ignaciuk, Base-stock distributed inventory management in continuous-review logistic systems-control system perspective, 2017 22nd International Conference on Methods and Models in Automation and Robotics, (2017), 1027–1032. Available from: https://ieeexplore.ieee.org/abstract/document/8046971
15. Y. Sun, C. Liang, S. Sutherland, et al., Modeling player decisions in a supply chain game, 2016 IEEE Conference on Computational Intelligence and Games (CIG), (2016), 1–8. Available from: https://ieeexplore.ieee.org/abstract/document/7860444
16. Giri, C.; Jain, S.; Zeng, X.; Bruniaux, P. A Detailed Review of Artificial Intelligence Applied in the Fashion and Apparel Industry.IEEE Access 2019,7, 95376–95396. [CrossRef]

APPENDIX

1. Machine Learning Algorithms

```
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report

# Instagram Profile Info
profiles = pd.DataFrame({
    "Instaprofile": [
        "absolutefashionbanaras", "lazree_sarees", "fabricity.co_",
        "fabricforever", "foreversilks"
    ],
    "Followers": [18700, 237000, 171000, 669000, 557000],
    "PostCount": [8, 11, 10, 10, 14]
})

# Likes and Shares Data
likes = pd.DataFrame({
    "Instaprofile": profiles["Instaprofile"],
    "Georgette": [1922, 7601, 1069, 5086, 102],
    "pure silk": [379, 1351, 133, 0, 1363],
    "silk": [0, 10201, 0, 136, 0],
    "crepe": [0, 250, 1006, 198, 4386],
    "net": [0, 0, 0, 5368, 1525]
})
shares = pd.DataFrame({
    "Instaprofile": profiles["Instaprofile"],
    "Georgette": [327, 1661, 207, 1502, 31],
    "pure silk": [72, 539, 18, 0, 375],
    "silk": [0, 3911, 0, 10, 0],
    "crepe": [0, 58, 173, 15, 854],
    "net": [0, 0, 0, 1259, 330]
})

# Long Format for Engagement Rate
likes_long = likes.melt(id_vars="Instaprofile", var_name="Material",
value_name="Likes")
shares_long = shares.melt(id_vars="Instaprofile", var_name="Material",
value_name="Shares")
eng = likes_long.merge(shares_long, on=["Instaprofile",
"Material"]).merge(
    profiles[["Instaprofile", "Followers"]], on="Instaprofile")
eng["EngRate"] = (eng["Likes"] + eng["Shares"]) / eng["Followers"]
```

```
# Mean Engagement Rate per Material
eng_material = (
    eng.groupby("Material")
       .apply(lambda g: (g["Likes"] + g["Shares"]).sum() /
g["Followers"].sum())
       .rename("EngagementRate")
       .reset_index()
)

# Inventory Data
inventory_data = pd.DataFrame({
    'Material': ['Georgette', 'crepe', 'net', 'pure silk', 'silk'],
    'UnitCost': [43968, 48951, 236774, 48839, 231157],
    'UnitPrice': [53617, 59654, 281974, 60270, 285183],
    'InboundInventory': [3570, 3260, 3560, 3600, 1985],
    'UnitsSold': [3465, 3096, 3462, 3538, 1965],
    'Deadstock': [105, 164, 98, 62, 20],
    'DemandTime': [5, 5, 5, 5, 5]
})

# Merge & Feature Engineering
df = pd.merge(inventory_data, eng_material, on="Material")
df["SellThroughRate"] = df["UnitsSold"] / df["InboundInventory"]
df["DeadstockRatio"] = df["Deadstock"] / df["InboundInventory"]
eng_threshold = df["EngagementRate"].median()

# Adjusted Optimization Criteria for balance
df["NeedsOptimization"] = ((df["DeadstockRatio"] > 0.02) &
(df["EngagementRate"] > eng_threshold)).astype(int)

# Feature Matrix and Labels
X = df[["UnitCost", "UnitPrice", "SellThroughRate", "DeadstockRatio",
"EngagementRate"]]
y = df["NeedsOptimization"]

# Train Models
lr = LogisticRegression()
lr.fit(X, y)
rf = RandomForestClassifier(random_state=42)
rf.fit(X, y)

# Predictions
df["LR_Prediction"] = lr.predict(X)
df["RF_Prediction"] = rf.predict(X)

# Reports
print("Logistic Regression Report:\n", classification_report(y,
df["LR_Prediction"]))
```

```
print("\nRandom Forest Report:\n", classification_report(y,
df["RF_Prediction"]))

# Feature Importance
importances = pd.Series(rf.feature_importances_,
index=X.columns).sort_values(ascending=False)
print("\nFeature Importance:\n", importances)

# Final Results
print("\nFinal Classification:\n", df[["Material", "NeedsOptimization",
"LR_Prediction", "RF_Prediction"]])
```

2. Python Code for Correlation Graph and Matrix

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Social Media Data
social_data = pd.DataFrame({
    'Material': ['Georgette', 'crepe', 'net', 'pure silk', 'silk'],
    'Likes': [15982, 5546, 6893, 3092, 10337],
    'Post Count': [18, 8, 11, 10, 8],
    'Mentions': [3789, 1040, 1589, 997, 3921]
})

# Step 2: Inventory and Sales Data
inventory_data = pd.DataFrame({
    'Material': ['Georgette', 'crepe', 'net', 'pure silk', 'silk'],
    'UnitCost': [43968, 48951, 236774, 48839, 231157],
    'UnitPrice': [53617, 59654, 281974, 60270, 285183],
    'InboundInventory': [3570, 3260, 3560, 3600, 1985],
    'UnitsSold': [3465, 3096, 3462, 3538, 1965],
    'Deadstock': [105, 164, 98, 62, 20]
})

# Step 3: Calculate Demand Score
social_data['Demand Score'] = (social_data['Likes'] +
social_data['Mentions']) / social_data['Post Count']

# Step 4: Merge datasets
df = pd.merge(social_data, inventory_data, on='Material')

# Step 5: Calculate rates
df['Sell-through Rate'] = df['UnitsSold'] / df['InboundInventory']
df['Deadstock Rate'] = df['Deadstock'] / df['InboundInventory']

# Step 6: Correlation matrix for selected features
```

```
selected_corr = df[['Demand Score', 'Sell-through Rate', 'Deadstock
Rate']].corr()

print("📊 Correlation Matrix (Selected Columns):")
print(selected_corr)

# Step 7: Scatter plot: Demand Score vs Sell-through Rate
plt.figure(figsize=(8, 5))
plt.scatter(df['Demand Score'], df['Sell-through Rate'], color='green')
for i in range(len(df)):
    plt.text(df['Demand Score'][i], df['Sell-through Rate'][i],
df['Material'][i], fontsize=9, ha='right')
plt.title('Demand Score vs Sell-through Rate')
plt.xlabel('Demand Score')
plt.ylabel('Sell-through Rate')
plt.grid(True)
plt.tight_layout()
plt.show()

# Step 8: Scatter plot: Demand Score vs Deadstock Rate
plt.figure(figsize=(8, 5))
plt.scatter(df['Demand Score'], df['Deadstock Rate'], color='red')
for i in range(len(df)):
    plt.text(df['Demand Score'][i], df['Deadstock Rate'][i],
df['Material'][i], fontsize=9, ha='right')
plt.title('Demand Score vs Deadstock Rate')
plt.xlabel('Demand Score')
plt.ylabel('Deadstock Rate')
plt.grid(True)
plt.tight_layout()
plt.show()

# Step 9: Correlation heatmap (boxed)
plt.figure(figsize=(6, 5))
sns.heatmap(
    selected_corr,
    annot=True,
    fmt=".2f",
    cmap="coolwarm",
    square=True,
    linewidths=0.5,
    cbar_kws={"shrink": 0.8}
)
plt.title("Correlation Matrix Heatmap (Selected)", fontsize=14)
plt.tight_layout()
plt.show()
```

Production Planning and Dead Stock Management Using Social Media Scraping.pdf

Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:97336076

Submission Date

May 23, 2025, 11:18 AM GMT+5:30

Download Date

May 23, 2025, 11:19 AM GMT+5:30

File Name

Production Planning and Dead Stock Management Using Social Media Scraping.pdf

File Size

4.4 MB

37 Pages

7,334 Words

43,664 Characters

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Match Groups

24 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
1% Publications
4% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Acceptance of Abstract for 3rd International Conference on Recent Trends in Mathematical Sciences (ICRTMS-2025)

1 message

ICRTMS2025 <icrtms25hgp@gmail.com>
To: Nancy Jain <nancyjain455@gmail.com>

Sun, Apr 13, 2025 at 3:39 PM

Dear Nancy Jain
I hope you are doing well.

We are pleased to inform you that the Conference Committee reviewed your abstract titled **"Production Planning and Dead Stock Management Using Social Media Scraping**" and has approved for presentation at "**3rd International Conference on Recent Trends in Mathematical Sciences (ICRTMS- 2025)"** scheduled to be held on **10th** – **11th May**, **2025** at **Himachal Pradesh University, Shimla**, H. P., India in Hybrid mode.

We believe that your presentation will make a valuable contribution to the conference. Your Paper ID is **ICRTMS_172**

We request you to **fill the registration form,** if not done already, and mail your f**ull length paper in PDF format** latest by **25th April, 2025.**
Please feel free to contact us for any queries.
To register, please fill out the Google Form available at the link:
https://forms.gle/X1qh8EtQetFBoXLe9

Lodging Arrangement
The organizing committee of ICRTMS-2025 makes arrangements for the stay of participants in nearby guest houses and hotels. The participants are free to exercise their choice about their stay for which they have to immediately contact the concerned guest house or hotel. **The participants are requested to book their accommodation by the end of March, 2025 as in the months of May and June there is tourist season in Shimla.**

Hotel City Inn: Situated at Lower Summer Hill and is about 500 meters from the venue.
Tariff: Rs. 560/- per person in double or triple sharing rooms.
Contact Details: +91-9418487752
Hotel Green View: Situated at Sangti and is about 1 km from the venue.
Tariff: Rs. 1000/- per person in double or triple sharing room with Balcony and Rs. 750/- per person in double or triple sharing room without Balcony.
Contact Details: **+91-98166-87459, +91-70186-26662, +91-78071-86043**
Hotel Ganga Palace: Situated at Summer Hill, Shimla and is about 100 meters from the venue.
Tariff: Rs. 3200/- per room (2 persons allowed), extra bed available (total 3 persons).
Group booking: Rs 900/- per person (4 persons per room)
Manager: Divesh Rathore
Contact Details: **+91-8262858998**

Thank you for your contribution to the conference.

On behalf of organizing committee

Dr. Neetu Dhiman
Convener
ICRTMS- 2025
Contact-+91-7018451738
Conference Website: https://icrtms25.hgp.org.in

Acceptance of Abstract for 3rd International Conference on Recent Trends in Mathematical Sciences (ICRTMS-2025)

1 message

ICRTMS2025 <icrtms25hgp@gmail.com> Sun, 27 Apr 2025 at 7:46 pm

To: Ishika Bansal <ishikabansal1931@gmail.com>

Dear Ishika Bansal
I hope you are doing well.

We are pleased to inform you that the Conference Committee reviewed your abstract titled **"Production Planning and Dead Stock Management Using Social Media Scraping"** and has approved for presentation at "**3rd International Conference on Recent Trends in Mathematical Sciences (ICRTMS- 2025)"** scheduled to be held on **10th – 11th May, 2025** at **Himachal Pradesh University, Shimla**, H. P., India in Hybrid mode.

We believe that your presentation will make a valuable contribution to the conference. Your Paper ID is **ICRTMS_172**

We request you to **fill the registration form,** if not done already, and mail your f**ull length paper in PDF format** latest by **25th April, 2025.**
Please feel free to contact us for any queries.
To register, please fill out the Google Form available at the link:
https://forms.gle/X1qh8EtQetFBoXLe9

Lodging Arrangement
The organizing committee of ICRTMS-2025 makes arrangements for the stay of participants in nearby guest houses and hotels. The participants are free to exercise their choice about their stay for which they have to immediately contact the concerned guest house or hotel. **The participants are requested to book their accommodation by the end of March, 2025 as in the months of May and June there is tourist season in Shimla.**

Hotel City Inn: Situated at Lower Summer Hill and is about 500 meters from the venue.
Tariff: Rs. 560/- per person in double or triple sharing rooms.
Contact Details: +91-9418487752
Hotel Green View: Situated at Sangti and is about 1 km from the venue.
Tariff: Rs. 1000/- per person in double or triple sharing room with Balcony and Rs. 750/- per person in double or triple sharing room without Balcony.
Contact Details: **+91-98166-87459, +91-70186-26662, +91-78071-86043**
Hotel Ganga Palace: Situated at Summer Hill, Shimla and is about 100 meters from the venue.
Tariff: Rs. 3200/- per room (2 persons allowed), extra bed available (total 3 persons).
Group booking: Rs 900/- per person (4 persons per room)
Manager: Divesh Rathore
Contact Details: **+91-8262858998**

Thank you for your contribution to the conference.

On behalf of organizing committee

Dr. Neetu Dhiman
Convener
ICRTMS- 2025
Contact-+91-7018451738
Conference Website: https://icrtms25.hgp.org.in

HIMACHAL GANITA PARISHAD

(REGISTERED UNDER H.P. SOCIETIES REGISTRATION ACT 2006)

3rd INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MATHEMATICAL SCIENCES (ICRTMS-2025)

10th-11th May, 2025

CERTIFICATE OF APPRECIATION

This is to certify that **Ms. Nancy Jain, UG/PG Student, Delhi Technological University** has presented a research paper entitled **Production Planning and Dead Stock Management Using Social Media Scraping** in **3rd International Conference on Recent Trends in Mathematical Sciences (ICRTMS-2025)** organized by the **Himachal Ganita Parishad (HGP) at Himachal Pradesh University, Shimla** on **10th-11th May, 2025.**

Dr. Kamalendra Kumar	**Dr. Neetu Dhiman**	**Dr. Shalini Gupta**
Co-Convener	Convener	President (HGP)

HIMACHAL GANITA PARISHAD

(REGISTERED UNDER H.P. SOCIETIES REGISTRATION ACT 2006)

3rd INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MATHEMATICAL SCIENCES (ICRTMS-2025)

10th-11th May, 2025

CERTIFICATE OF APPRECIATION

This is to certify that **Ms. ISHIKA BANSAL, UG/PG Student, Delhi Technological University** has presented a research paper entitled **Production Planning and Dead Stock Management Using Social Media Scraping** in **3rd International Conference on Recent Trends in Mathematical Sciences (ICRTMS-2025)** organized by the **Himachal Ganita Parishad (HGP) at Himachal Pradesh University, Shimla** on **10th-11th May, 2025.**

Dr. Kamalendra Kumar	**Dr. Neetu Dhiman**	**Dr. Shalini Gupta**
Co-Convener	Convener	President (HGP)