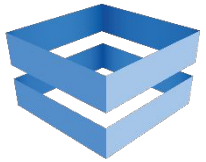
Mid Term Report

AI-Powered
Product Classification

Rujul Dwivedi



latentview

Actionable Insights • Accurate Decisions

- Latentview Analytics

  - ***Specializes in AI-driven data analytics solutions.***
  - ***Provides end-to-end analytics to global industry leaders.***
  - ***Focus on business insights, automation, and AI innovation.***

Project Scope

Challenge

Manual product classification is error-prone and inefficient.

Goal

Develop an AI-powered system to automate product categorization.

Actionable Insights • Accurate Decisions

Key Objectives

1. **Automate product categorization.**
2. **Improve classification accuracy.**
3. **Reduce manual effort and errors.**
4. **Gain market insights through classification analysis.**

LatentView
Actionable Insights • Accurate Decisions

Methodology & Approach

Data Collection

Gathered data from Amazon datasets and preprocessed.

Exploratory Data Analysis (EDA)

Visualized class distribution and correlations.

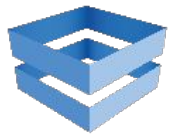
Model Development

Implemented LSTM as baseline model and explored TF-IDF vectorizations.

Optimization

Addressed data imbalance with undersampling/oversampling strategies.

Results



latentview

Actionable Insights • Accurate Decisions

Scores

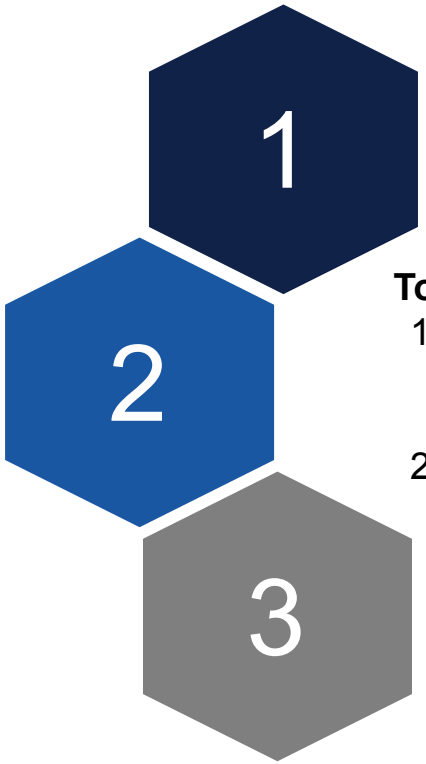
89% accuracy in predicting the test set titles to their correct category

Increasing epochs from 5 to 10 (but causes overfitting)

89% ~ > 92%

Actionable Insights • Accurate Decisions

Key Findings



Accuracy Trends:

1. Increasing epochs caused overfitting (>92% but poor test performance).
2. Larger dataset fractions (0.1 to 0.3) improved generalization.

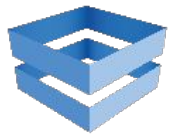
Top Errors & Misclassifications:

1. Skewness causing some categories to be frequently misclassified.
2. Identified top 3 misclassified categories per class.

EDA Insights:

1. Strong correlation between missing titles and empty descriptions.
2. Word clouds & class histograms provided feature insights.

Future Work



LatentView

Actionable Insights • Accurate Decisions

Second Half

01 Enhancing feature extraction with deep learning-based embeddings.

02 Hyperparameter tuning for transformer-based models.

03 Building GenAI models for direct text-based classification.

04 Real-world validation on larger datasets & high-performance servers.

LatentView
Actionable Insights • Accurate Decisions

latentview

Actionable Insights • Accurate Decisions

Thank You

Socials

www.iitgoa.ac.in
www.latentview.com