Executive Report: Women's Accessories Image Classification Model

Image Classification Model

Contents

Prepared By:

Thapelo Mogashoa   600298
Janco Nieuwoudt   600175
Dean Jacobus Andreas van Zyl   600367

Introduction

Accessory Avenue operates as an independent e-commerce platform which shows interest in growing its seller database after separating from its parent organization. Accessory Avenue hired our team to create an image classification model which improves how new sellers join the platform. The model functions to sort product images from prospective sellers between women's accessories which include jewelry and handbags as well as scarves and hats and belts and watches and non-accessory items that consist of dresses and shirts and shoes and non-product pictures. This goal will boost the platform's seller application review speed to maintain relevant products only on Accessory Avenue.

The document presents details about the visual category classification model creation process including performance indicators and far-reaching effects from the Jupyter Notebook (Product_Image_Classification_Notebook.ipynb) analytic evaluation.

1. Business Understanding

Accessory Inc is a new online e-commerce platform, that's focuses on women's accessories. The platform is currently facing challenges onboarding new sellers, while maintaining their current product catalog. Accessories Inc separated from its parent company, giving it a massive opportunity to redefine their brand identity. They currently want to sell women's accessories such as jewellery, handbags, scarves, hats, belts, and watches. However they are currently manually vetting each individual seller, which is time consuming and prone to human error, and is currently a bottleneck in their operations. They worry that their inefficiency might be driving their potential sellers to their competitors, while leaving them with an incomplete catalogue.

They've hired us to create an image classification model to help automated the initial screen process. This will help the reduce their labour costs as they will no longer need to hire someone to constantly check the images on the hour. This will allow them to scale the business efficiently, as they will be able to vet thousands of people at a time without worrying about inaccuracies, and the onboarding processes for the sellers will be faster than ever before.

Key stakeholders include:
- **Accessory Avenue Management**: Seeks operational efficiency and brand consistency to drive platform growth.
- **Sellers**: Looking forward to a streamlined onboarding process.
- **Customers**: Expecting a high quality, focused Catalogue of women's accessories
- **Data Science Team**: Responsible for developing, deploying, and maintaining the model to meet business needs.

The project will be deemed successful by analysing the accuracy of the image classification model.

2. Project Overview

2.1 Objective

Creating a model that can easily classify images as women's accessories.

- **Women's Accessories** (labeled as 1): Includes items such as jewelry, handbags, scarves, hats, belts, and watches.
- **Non-Women's Accessories** (labeled as 0): Includes other clothing items (e.g., dresses, shirts, shoes) or non-product images.

The model will help Accessories Inc classify their images.

2.2 Dataset

The model was developed using the **Fashion Products Images** dataset from Kaggle, which includes:

- A CSV file (styles.csv) containing metadata for various fashion products, including product ID, gender, master category, subcategory, article type, and image file names.
- A collection of corresponding product images stored in an images directory.

Key dataset characteristics:

- **Total Entries**: 44,441 products.
- **Women's Accessories**: Identified by filtering for products with gender = Women and masterCategory in ['Accessories', 'Personal Care'], resulting in 7,130 images.
- **Non-Women's Accessories**: All other products, including men's items, other women's clothing, and other items.
- **Balanced Subset for Training**: To address class imbalance, the training dataset was limited to 10,000 images (5,000 women's accessories and 5,000 non-accessories).
- **Testing Dataset**: 4,260 images (2,130 women's accessories and 2,130 non-accessories) were reserved for testing, ensuring no overlap with the training set.

2.3 Methodology

The Jupyter Notebook outlines the following steps:

1. **Data Preprocessing**:
   - Loaded and explored the dataset using Pandas to understand product categories and distributions.
   - Filtered for women's accessories based on gender and masterCategory.
   - Created a balanced dataset by sampling equal numbers of positive (women's accessories) and negative (non-accessories) images.
   - Processed images using OpenCV, resizing them to a uniform size (using Lanczos interpolation) and normalizing pixel values.
2. **Model Development**:
   - A convolutional neural network (CNN) was implemented using TensorFlow/Keras.
   - The model was trained for 20 epochs on the balanced training set of 10,000 images.
   - Validation was performed during training to monitor performance on a separate validation set.
3. **Evaluation**:
   - The trained model was evaluated on the test set (4,260 images) using metrics such as accuracy, precision, recall, and F1-score.
   - Training and validation accuracy were visualized to assess model learning and potential overfitting.
4. **Model Deployment**:
   - The final model was saved as imageprediction.h5 for integration into Accessory Avenue's seller application pipeline.

3. Statistical Performance

The model's performance on the test set (4,260 images) is summarized below:

| Metric	Value	Interpretation
Accuracy	0.9167	91.67% of images were correctly classified as women's accessories or not.
Precision	0.6657	66.57% of images predicted as women's accessories were actually accessories.
Recall	0.9652	96.52% of actual women's accessories were correctly identified.
F1-Score	0.7880	Harmonic mean of precision and recall, indicating a balanced performance.

3.1 Analysis of Results

- **High Accuracy (91.67%)**: The Model Does really well on the training data scoring an impressive accuracy 91.67%
- **High Recall (96.52%)**: The model excels at identifying women's accessories, with very low false negatives . this ensures that most valid seller submissions are not rejected.
- **Moderate Precision (66.57%)**: This indicates that non accessory items are wrongly classified as women accessories, so accessories inc would need to have additional validation steps in their procedure, to account the false positives
- **F1-Score (78.80%)** : The F1 Score indicates the balance between eh Recall and the precision at 78.8 percent is good but can be improved.

3.2 Training Insights

- The training accuracy and validation accuracy plots, show us some useful information
  - **Convergence**: Both training and validation accuracy increase steadily over the 20 epochs, indicating effective learning.
  - **No Significant Overfitting**: The validation accuracy closely tracks the training accuracy, suggesting the model does well on unseen data, indicating that there is no overfitting

A Graph showing the training Accuracy and the Validation Accuracy per epoch
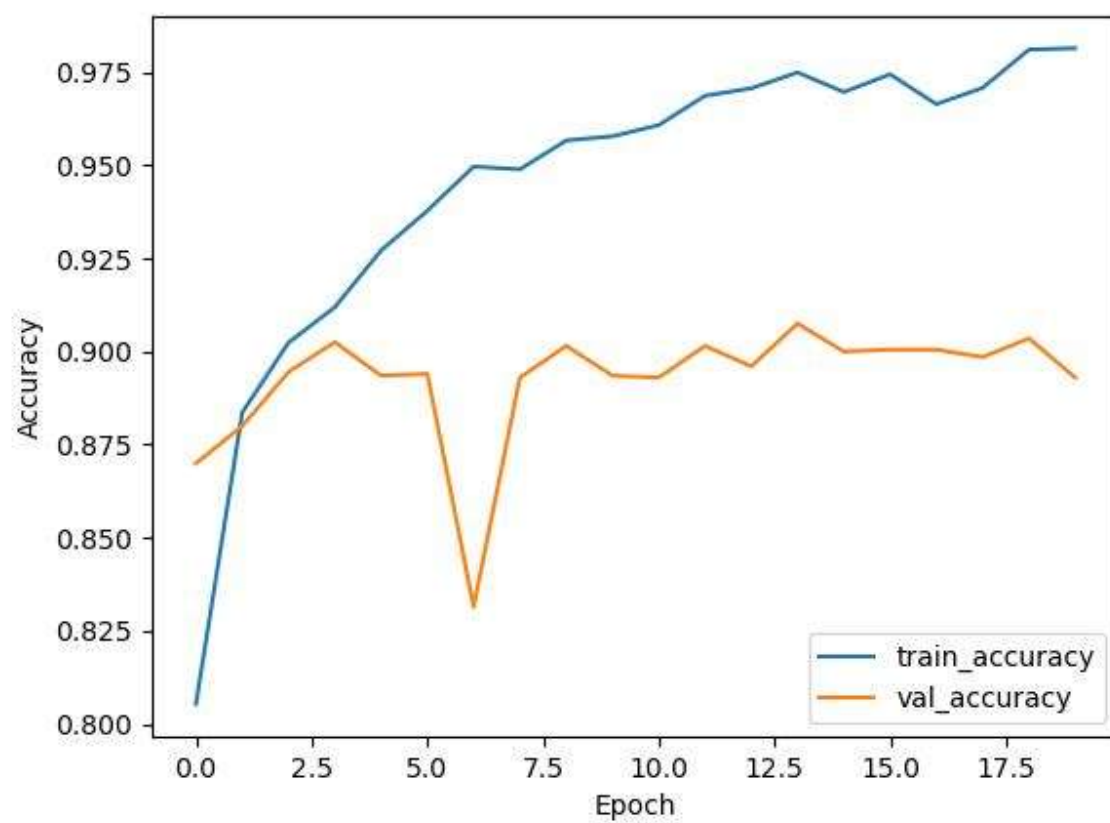
4. Business Implications

4.1 Benefits for Accessory Avenue

- **Automation of Seller Application Filtering**:

  - The model system executes the first step of product image screening thereby requiring less time from employees to examine merchant applications.

  - Through its 91.67% accuracy rate and 96.52% recall rate the model effectively identifies most genuine women's accessory submissions thereby helping Accessory Avenue welcome suitable new sellers.
- **Improved User Experience**:
  - By ensuring that only women's accessories are listed, the model helps maintain a curated product catalog, enhancing customer trust and satisfaction.
- **Scalability**:
  - The model can process large volumes of image submissions, enabling Accessory Avenue to scale its seller base efficiently as the platform grows.

4.2 Limitations and Considerations

- **Moderate Precision**:
  - The 66.57% precision demonstrates significant incorrect classification problems when non-accessories get labeled as accessories. The 66.57% precision leads to possible extra expenses because it may necessitate secondary manual review for certain applications.
  - Accessible Avenue needs to evaluate the automation versus human review costs by setting a threshold for uncertain predictions that would need manual verification.
- **Dataset Bias**:
  - The dataset originates from a parent company that markets many products so it might lack full representation of the women's accessories which will be offered by Accessory Avenue. The upcoming model version needs to include genuine market data from Accessory Avenue's prospective customer base.
- **Generalization**:
  - The model was trained on a subset of the Kaggle dataset, which may not capture The model received training data from a part of the Kaggle dataset but the dataset likely failed to include the complete range of seller-submitted images including various product types and different image conditions and backgrounds. The recommendation calls for testing on genuine seller contents from the real world.

5. Recommendations

To maximize the model's effectiveness and align it with Accessory Avenue's business goals, we propose the following:

1. **Enhance Precision**:
   - Model performance can be enhanced through two methods: conduct threshold changes in combination with data training to minimize incorrect positive outcomes.
   - The model's results will benefit from testing of advanced network architectures including ResNet and EfficientNet alongside transfer learning implementation for better features extraction and classification quality.
2. **Integrate Confidence Scoring**:
   - Through predictive probability output the system enables Accessory Avenue to select high-scored predictions and send low-scored ones to manual examination.
3. **Expand Dataset**:
   - The team will gather a proprietary image collection of women's accessories tailored for Accessory Avenue's business segment to enhance model precision and adaptation abilities.
   - The system includes diverse realistic image conditions that match the actual submission standards of sellers.
4. **Deploy and Monitor**:
   - The saved model (imageprediction.h5) will be integrated with Accessory Avenue's application process through either a web interface or API deployment.
   - The model performance should be tracked with current data from the real world to enable site-wide adaptations whenever new seller behaviors or product trends emerge.
5. **User Feedback Loop**:
   - Sellers whose applications get rejected can use the established feedback system to present corrected images and appeal their decisions. Filings sent to the model will help improve its precision for future evaluations.

6. Conclusion

The model built for image classification delivers 91.67% accuracy and 96.52% recall which makes it ideal for automated seller application screening at Accessory Avenue. The accurate accessory detection capability of the model aids Accessory Avenue in its goal to curate the catalog while facilitating new seller onboarding. The current model performance delivers a decent base for operations expansion through its moderate precision level (66.57%) but can still be improved. The proposed enhancements will allow Accessory Avenue to maximize its model while achieving long-term business goals for continuous high-quality seller onboarding efficiency. Our data science team handles all inquiries regarding additional information.

For any questions or further details, please contact our data science team.

Additional Info

Git hub : https://github.com/CantCode29/MLG382ProjectFolder

Dash app : https://mlg382projectfolder-1-wgce.onrender.com/