Flipkart

GRID 2.0

Fashion Intelligence Systems

Team Name   : FlipBlitz

Institute Name: Birla Institute of
Technology & Science (BITS), Pilani

1

Team members details

Team Name	FlipBlitz		
Institute Name	Birla Institute of Technology and Science (BITS), Pilani		
Team Members >	1 (Leader)	2	3
Name	Pratyush Goel	Aditya Vishwakarma	Shrey Shah
Batch	2022	2022	2022

Glossary

- KSV : Keyword Search Volume.
- TF-IDF (Term Frequency - Inverse Document Frequency): is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.
- EMA : Exponential Moving Average.
- SARIMA : Seasonal Autoregressive Integrated Moving Average
- CNN : Convolutional Neural Network

Use-cases

Priority (p0):

- The retailer can view **current trending** apparels in sorted order, based on the "trend score".
- The retailer can view the **upcoming trend** of apparels in a rank-wise manner and customize their fashion portfolio.

Priority (p1):

- The retailer can filter trend wrt fields like brand, fabric, pattern etc to gain more insight of ongoing and upcoming trends.

Priority (p2):

- The website will be integrated with e-commerce websites to allow the user to purchase trending products.

Proposed approach

The solution aims to predict ongoing and upcoming fashion trends and enable fashion retailer to customize their portfolio. The solution is divided into 4 phases.

1. Data scraping from fashion e-commerce websites.
2. Predicting the current trending products from the scraped data and sort them based on the trend score.
3. Predicting the future trend on these products by collecting time-series data from past trend and current blogs (which acts as influencers for future trends).
4. An interactive, easy to use UI that enables the retailer view trend and customize their portfolio.

Here is a Design-brief of the website:

https://pratyushgoel102666.invisionapp.com/freehand/Fashion-website-SNxqylJli

PHASE 1: Web Scraping

1) In the first step, the data is scraped from online, fashion-centric, e-commerce website using *Octoparse* to automate data extraction process.

2) The scraped data is of two types: -

   a) Data with images as well as a description about the product

   b) Data with only images which lack proper descriptions.

3) For a), NLP methods (using libraries like *SpaCy* and *Scikit-learn)* would be applied on the 'description' of the products, to extract **fashion attributes** (pertaining to)/of that product.

   For b), the **fashion attributes** are extracted from only the images, using a **CNN based fashion-attributes prediction model** trained on the *DeepFashion* dataset[1].

4) Finally, we have uniform data in a usable form where we have the url of the product, its image and most importantly their **fashion attributes.**

1 - Reference for attribute prediction model: https://arxiv.org/pdf/1705.06394.pdf

Flipkart | GRID 2.0

PHASE 2: Current Trend Prediction

- Current trend of any product will be calculated on the basis of **trend score** of each of its attribute.

- Trend score is calculated on the basis of **keyword search volume(KSV)** of each of the product's attributes.

Algorithm to calculate **trend score: -**

1. The trend score will be collected from *pytrends* library which uses **google trends API**.

2. The current trend score for each attribute is equal to the *Exponential moving average(EMA)* on the present day, calculated on KSV data from *past 3 months*.

- Each attribute of the product will be queried to google trends with respect to a **base attribute**, to get a relative trend scores.

3. After finding the EMA of each fashion attribute of a single product, the final trend score for that product is the **average mean over EMA of each of its fashion attributes**.

PHASE 3: Future Trend Prediction

- Using the past 5 year data from google trends, the **SARIMA** model will forecast the trends for next 3 months wrt each attribute of the apparel.
- From over *100 recent fashion blogs*, the attribute prediction model will find all the fashion attributes of apparel from the images.
- **Tf-idf/ normalized_frequency** will be calculated to score trend from the blogs.
- *Global_trend_score* =

  Average(**w1***trend_score(SARIMA) + **w2***trend_score(blog))

  Here w1 and w2 are parameters/weights,
- *Py3-pinterest* library will be used to forecast trend of attributes like colors using repin count.

Animal Prints on Fall/Winter 2018 and Spring 2019 Runways

Burberry | Tom Ford | Zimmermann | Versace

Chart 2: New-In Movement on Animal Print Across Seasonality

Flipkart | GRID 2.0

PHASE 4: Progressive Web Application

- **Elasticsearch** will be used to manage huge amount of data and achieve faster performance with search result for the search engine.
- **Appbase.io** will be used as a hosted search backend on top of Elasticsearch to make storing, querying and streaming data easily.
- **ReactiveSearch** and **React** libraries for frontend to handle UI rendering, query requests and manages response state.
- Finally, the complete end to end solution will be deployed on **Firebase**.
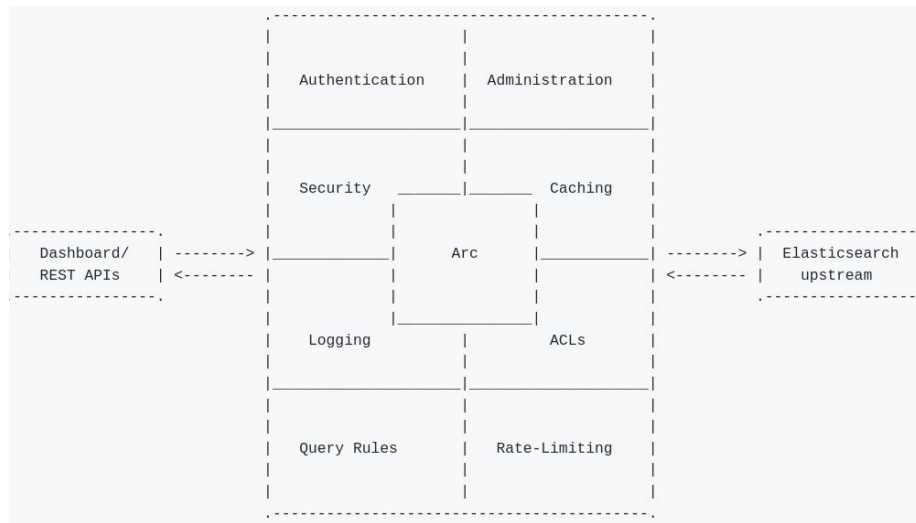
System Workflow for showing current and future trends
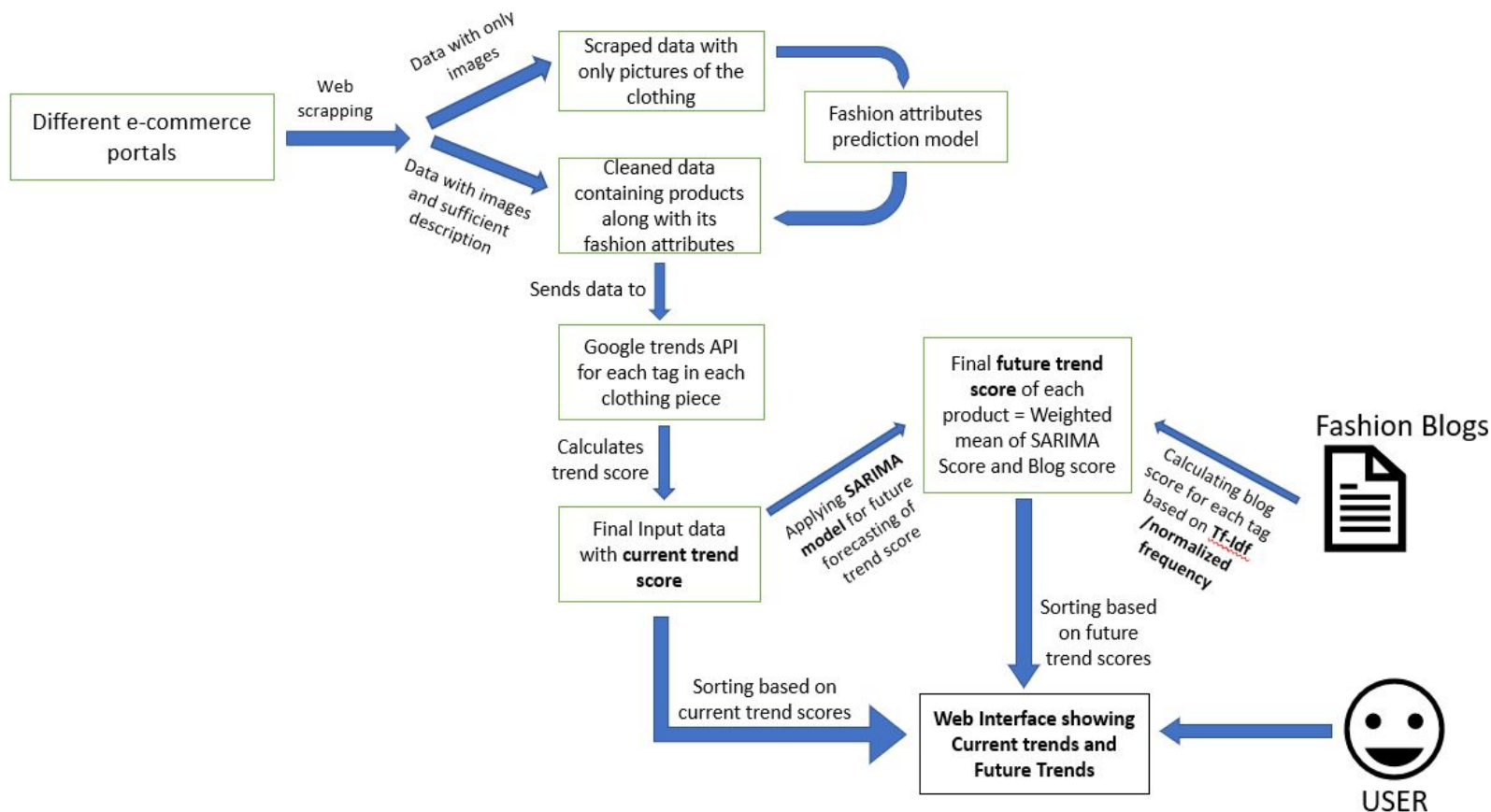
Design choices

PRO
- Fashion attributes are extracted from images when there is a lack of textual description.
- Compared to other Machine Learning models, SARIMA considers seasonal variation to predict global trend
- Using Elasticsearch instead of other database for complex search and quick analysis of large volumes of data.

CON
- Updates are expensive in Elasticsearch. Hence, analysis provided is not real time, but it will be periodically updated.
- User cannot get an absolute score about trendiness of an object.
- Experimental estimation of weights(w1,w2), to calculate trend score.

Additional Reading Material

- https://towardsdatascience.com/understanding-sarima-955fe217bc77
- https://github.com/appbaseio/arc (for web application)

Limitations

- Due to the unavailability of sales data in the public domain, peoples trend of purchasing the apparel are unaccounted.
- There is no definite mechanism for us to determine the accuracy of our model with apparel trend curve.
- As Elasticsearch only supports JSON format, no other text file is supported.
- Our solution is very data-intensive. To get a general idea of the upcoming trend for all apparels, we would require a huge number of recent blogs.

Future Scope

1. Addition of vertical scrollbar widgets corresponding to each attribute of the apparel. This will enable the fashion retailer to add his experience/vision as a score and rearrange the trending apparels.
2. Celebrities have a huge impact on the modern fashion trend. So, social media parameters (for eg Instagram likes) can be taken into account to calculate the trend score. This can be achieved using image retrieval techniques to find similar apparel present in the data and score them accordingly.

- **Reference:** https://wenxinxublog.wordpress.com/2017/01/06/deep-shopping/