

Search Space Reduction in Fashion E-Commerce: Applying NER and Classification for Better Search Outcomes

STUDENT

YASH RAJ SINGH
(231014984)

SUPERVISOR

RAHUL DUTTA



Agenda

- | | |
|----|---------------------------------|
| 1 | <u>Introduction</u> |
| 2 | <u>Motivation</u> |
| 3 | <u>Objectives</u> |
| 4 | <u>Methodology</u> |
| 5 | <u>Architecture Overview</u> |
| 6 | <u>Named Entity Recognition</u> |
| 7 | <u>Classification</u> |
| 8 | <u>Elastic Search</u> |
| 9 | <u>Results</u> |
| 10 | <u>Demo</u> |
| 11 | <u>Conclusion</u> |
| 12 | <u>Future Work</u> |



Introduction

- Fashion e-commerce: vast, competitive industry
- 2021: US\ \$553.1 billion , 2027: US\ \$1,222.32 billion - Mailmodo
- Catalogue Increasing
- Critical need for relevant search results
- Search Space is huge

Motivation

- Develop a system that helps in reducing the number of searches.
- Make the User experience less overwhelming.
- Reduce the load on real time resources



Objectives

- Improve search relevance using advanced techniques : Named Entity Recognition and Logistic Regression Classifier
- Focus on reducing volume of product on user catalogue
- Reduce irrelevant search results.

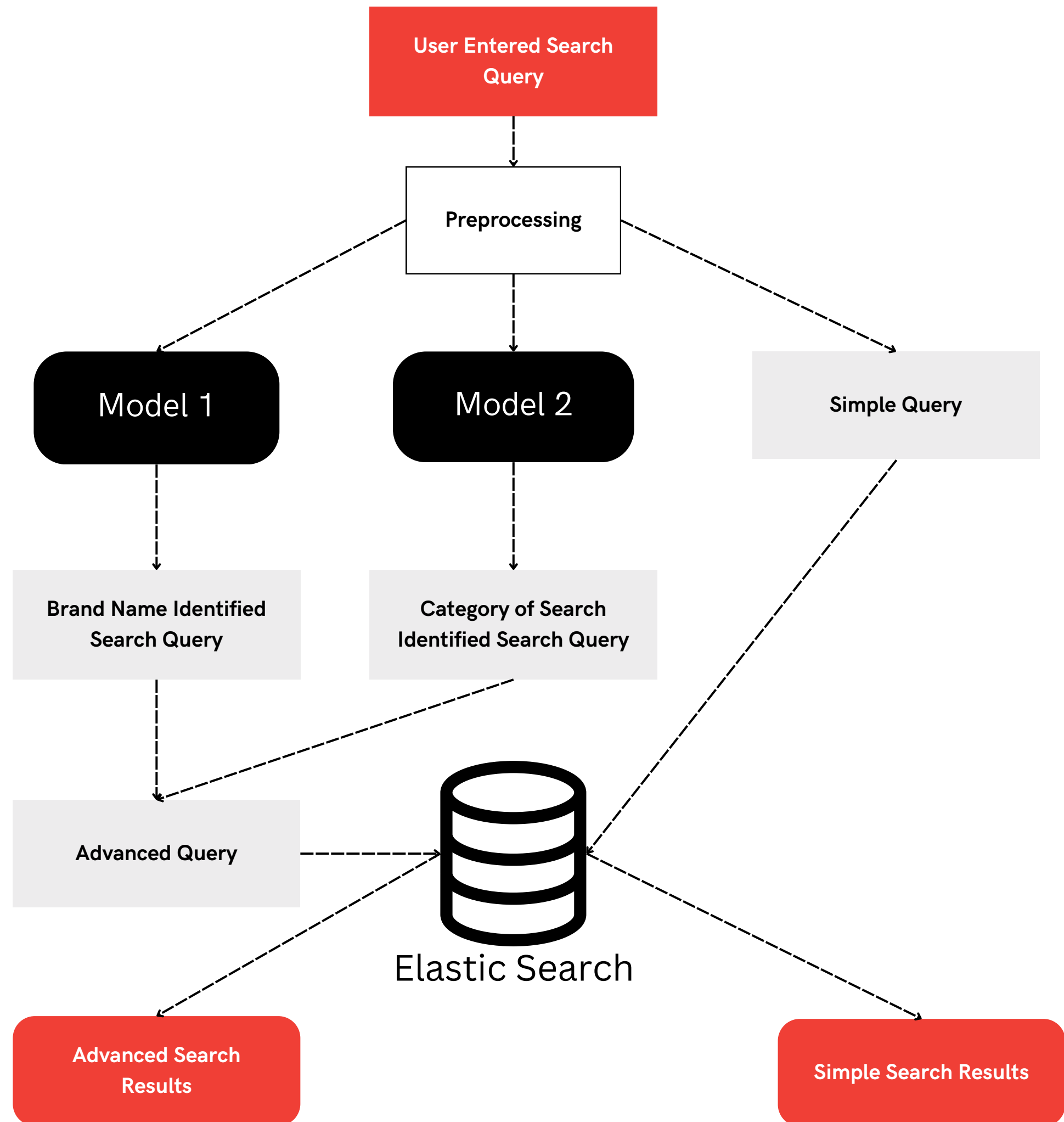


Methodology

- Data Collection - Myntra Fashion Dataset
 - Data Preprocessing
 - Named Entity Recognition : Brand Name Recognition from Search Query
 - Classification : Find category of product from Search Query
 - Integrate with Elasticsearch for Search functionality
 - Flask served web app for integration
 - Two search modes: Simple vs. Advanced
 - Use these Information to reduce the number of searches user sees.
-



Architecture Overview



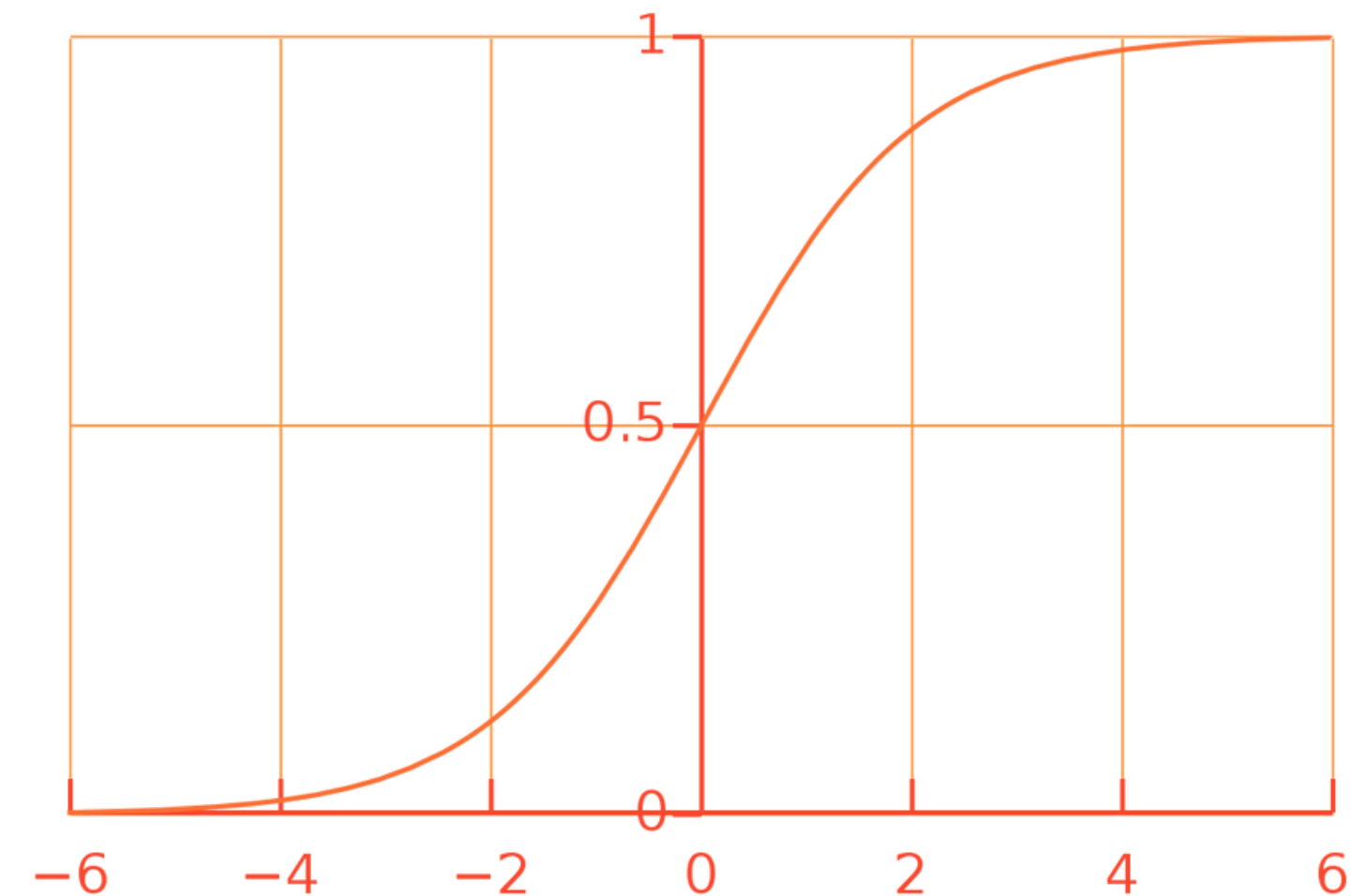
Model 1 - Named Entity Recognition - SPaCy

- Used SpaCy, a powerful NLP library, for NER.
- Trained a custom model with domain-specific data. i.e. Fashion data.
- Identified brand names are further passed into system flow.



Model 2 - Classifier - Logistic Regression

- Preparing the labelled data to identify the category of search query.
- Vectorize the features for better training.
- Trained the model on processed labelled data.



Search Engine - Elastic Search

- Serve as the search engine to handle and optimize search queries.
- Provide fast and scalable search capabilities across large datasets.
- Stores data in form of documents in indexes.



Ensembling..

- Simple Search
 - Query on Elastic Search Index with original search query.
- Advanced Search
 - The outputs from the two models are used to reduce the number of result.
 - Model 1 - Brand Name
 - Model 2 - Category
 - Query is updated with additional information from Models

```
if search_query:
    # Simple search query
    simple_search_body = {
        "query": {
            "match": {
                "description": search_query
            }
        },
        "from": offset,
        "size": RESULTS_PER_PAGE
    }
```

Simple Search Query

```
must_clauses = [{"match": {"description": search_query}}]
if identified_brand:
    must_clauses.append({"match": {"brand": identified_brand}})
if predicted_category:
    must_clauses.append({"match": {"category": predicted_category}})
advanced_search_body = {
    "query": {
        "bool": {
            "must": must_clauses
        }
    },
    "from": offset,
    "size": RESULTS_PER_PAGE
}
```

Advanced Search Query

Results

Results of Different metrics on
Named Entity Recognition

NER Model Metrics	
<i>Metric</i>	<i>Value</i>
Precision	0.8610
Recall	0.8539
F1 Score	0.8565
Accuracy	0.9863

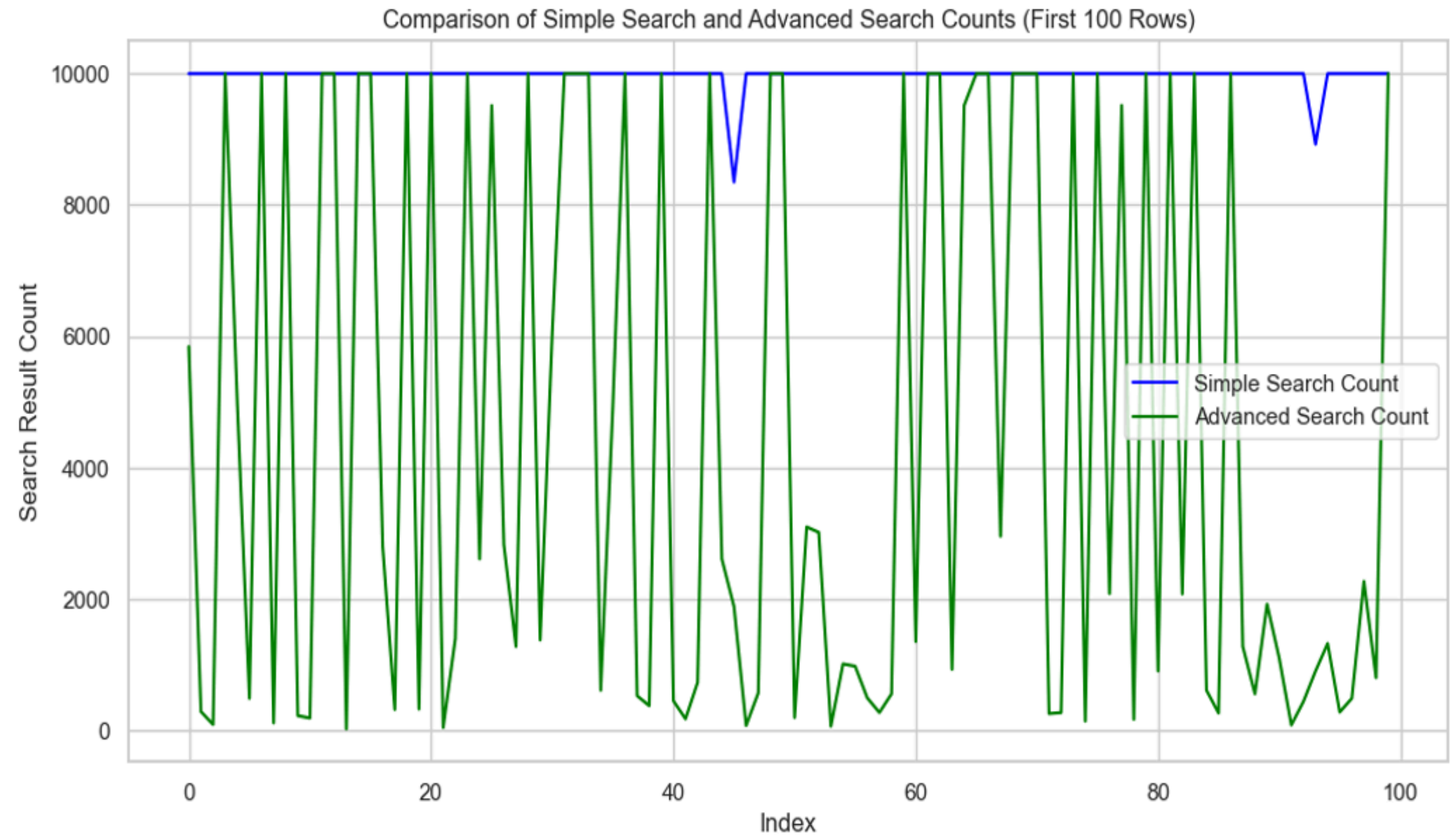
Results

Results of Different metrics on
Logistic Regression Classifier

<i>Metric</i>	<i>Logistic Regression</i>
Precision	0.9485
Recall	0.9464
F1 Score	0.9472
Accuracy	0.9653

Results

Graph of variation in number of search results for a sample of 100 searches



Percentage Reduction in Number of Search Results from Simple to Advanced Search Method



[104]

```
1 mean_reduction_percentage = df['reduction_percentage'].mean()  
2 mean_reduction_percentage
```

✓ 0.0s

...

54.16849034496216

Demo



Conclusion

- Using SpaCy's NER to accurately identify brand names and Logistic Regression to categorize queries, effectively narrowing down search results.
 - The reduction in irrelevant search results in the advanced search highlights the system's ability to focus on products that are more pertinent to users' needs.
 - This leads to a more satisfying shopping experience and has the potential to increase conversion rates.
-

Future Work

- Extend NER and classification models to support multiple languages.
 - Adapt the system for broader e-commerce categories beyond fashion apparel.
 - Implement mechanisms to gather and analyze user feedback on search results.
-

Thank You

