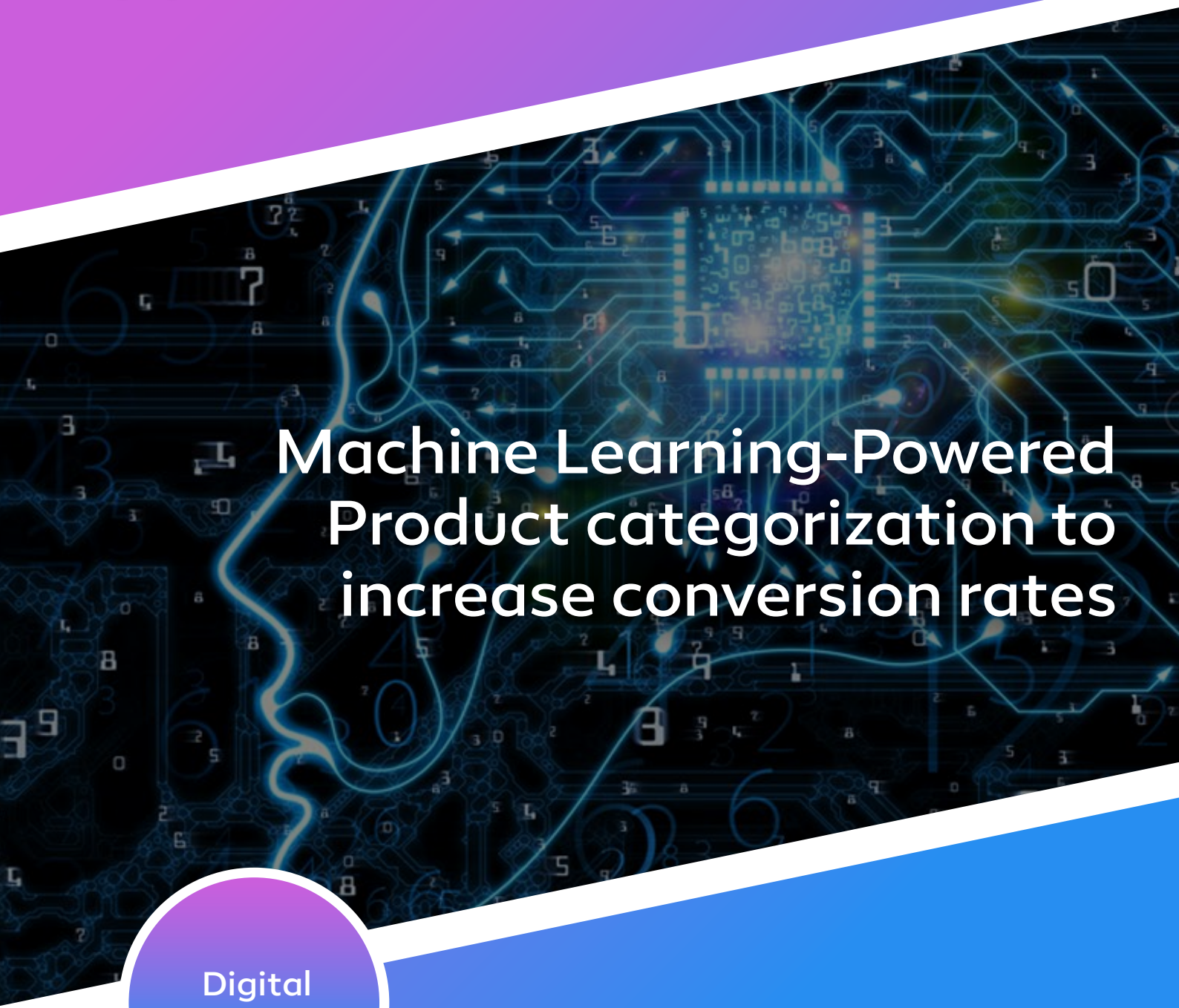
indium
Make Technology Work



Machine Learning-Powered Product categorization to increase conversion rates

Digital Services

Success Story

www.indiumsoftware.com

Client

The client is an AI-powered e-commerce aggregator website which delights customers by providing them with smart buying options.

Overview

As with any online retailer, product categorization is an indispensable functionality required to offer both text and text-free searches, recommendations and up-selling. While the client was capable of categorization on a retailer's website, it was experiencing insurmountable challenges across multiple websites. Indium Software leveraged its expertise in Advanced Machine Learning and Artificial Intelligence to devise a "name matching" model that overcame these obstacles improving the accuracy of the listings on the client's website.

Challenges

The product categorization is indispensable for e-commerce websites. It makes free-text searches faster and provides better user experience by highlighting top categories upfront.

While this categorization works well on retailer websites, it becomes an issue for e-commerce aggregators. The product categories are defined differently by different retailers for the same product. This creates a problem in assigning the same product from different retailers to the same categories due to which the quality of search results and user experience suffer.

There was a need for Advanced Machine Learning and Artificial Intelligence techniques to be deployed to solve some of the most complex problems the industry faces. In particular, this would be a case on how we solved the nagging problem of product categorization.

Business Challenges

Here is a summary of the steps involved and the methods used in each step of this process.

- Data Sampling
  - To get equal representation from each category.
  - Methods: Stratified Random Sampling.
- Pre Processing
  - Convert Text to a numeric representation.
  - Methods: TF-IDF, N grams, Stop words, stemming lemmatization.
- Model Training
  - Fit classification models to training data.
  - Naive Bayes, Support Vector Machines.

Business

Machine Learning

Domain

eCommerce

Tools

Python, Scikit Learn, Jupyter Notebook (for prototyping), PySpark, Pickling using Django/ Flask (for scaling and production)

Key Highlights

- The product categorisation model achieved a 75% accuracy in classifying the products
- 3% increase in conversion rates across all Product Categories, which led to a 20% increase in GMV

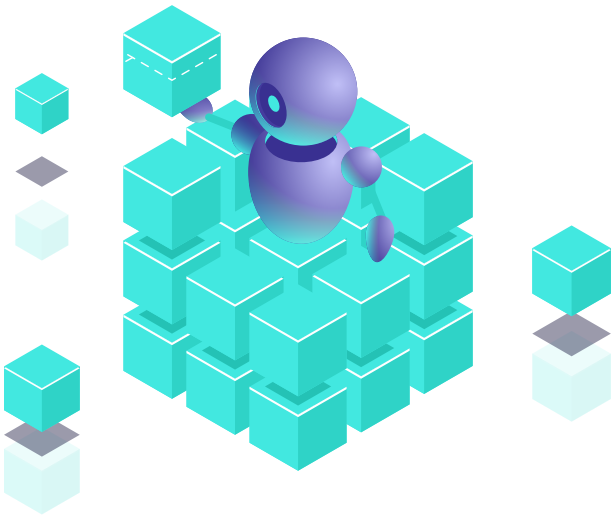
- Parameter Tuning
  - Find model parameters that give the highest accuracy.
  - Grid Search (Scikit-learn), Cross Validation.

- Model Nesting
  - Train separate models for different levels of hierarchical groups.
  - Group By methods, For-loops, Sub-setting.

- Production
  - Setting up infrastructure for the trained model to be utilized.
  - Pickling, Django, JavaScript.

Business Impact

- Model trained achieved an accuracy of 75% in accurately predicting categories for new products.
- Improved Product Categorization lead to superior indexing of products which directly contributed towards providing better search results.
- 3% increase in Conversion rates across all Product Categories which led to a 20% increase in GMV.



INDIA

Chennai | Bengaluru | Mumbai
Toll-free: 1800-123-1191

USA

Cupertino | Princeton
Toll-free: +1 888 207 5969

UK

London
+44 773 653 9098

SINGAPORE

+65 9630 7959

MALAYSIA

Kuala Lumpur
+60 (3) 2298 8465

General Inquiries
info@indiumsoftware.com

Sales Inquiries
sales@indiumsoftware.com