Hackathon
Subject : Your Next Purchase

Recommending the right product to each client for a marketing action

February 10th – 15th, 2025

Zongru SUN, Binong HAN, Jinsuh YOU, Julien DEHUT, Solène LEBOUCHER, Tonghan WEN

CentraleSupélec ESSEC BUSINESS SCHOOL eleven Strategy x Artificial Intelligence

A need to automate product recommendations

Our client, a player of the retail industry, would like to improve the performances of marketing actions

For now, they only have a manual product recommendation system generalized to all clients

They want to implement a recommender system based on AI, personalized to each client

- What products is each client most likely to buy?
- How can we leverage these insights for strategic recommendations?

An AI-based process which allows immediate product recommendation

Current manual process

Prospect comes to a shop or creates an account → Prospect becomes a client → First purchase anniversary → Client is at risk of churning

- Low customization (one-to-many)
- Lack of visibility on performance

AI-powered process

Prospect becomes a client → Automatic AI-based product recommendation → Client makes new purchases

- Immediate and automatic generation of product recommendation
- Consideration of the available stock

- Continuous metrics improvement
- Test and learn approach made possible

Objectives (Business-View)

Design an automatic personalized recommendation system to max long-run profit

Revenue (Increase) − Cost(Decrease)

# of client × # of transaction × Revenue / transaction

- Increase in Loyal Proportion
- Increase in transactions
- Consider gross profit when recomending products

Cost of stock + Cost of stock + Cost of stock

- Optimize recommendation considering stock info
- Optimize recommendation considering store place

$$\textbf{Potential ROI} = \frac{\text{Increase in profit in long-run view}}{\text{Cost of designing \& updating model} + \text{implementing marketing strategies}}$$

Objectives (Technical-View)

Design an automatic personalized recommendation system

Revenue (Increase) — Cost(Decrease)

# of client × # of transaction × Revenue / transaction

- User profile construction
- Product similarity calculation
- Complementary product recommendation
- Diversed & Novel recommendation

Cost of stock + Cost of transportation

- Product stock tracking
- Store & Client location tracking

Metrics: $$Recall@k = \frac{nb\ of\ products\ bought\ among\ the\ k\ recommendations}{nb\ of\ products\ bought\ by\ the\ customer}$$

Key Success Factors

	WHY?	HOW?
Enhance 'Loyal' customer experience	• Majority of clients are loyal • Generating most revenue	• Recommend most similar products with one more totally new different product • More transactions = more loyal
Reduce Churn Rate & Improve Retention	• Inactive - > loyal • Loyal clients base - > more transactions & more revenue	• Recommend based on previous transactions & similar customers
Increase Conversion Rate	• Attract potential customers who browse but don’t buy	• Use real-time recommendations (e.g., "X people bought this today") to encourage purchases
Increase Cross-selling & Upselling	• Maximize revenue and profit by recommending several products & higher-value products	• Market Basket analysis: find commonly purchased products, promotions & discounts • Rank products by profit margin
Enhanced Operational Efficiency	• Reduced stock cost and transportation cost	• Optimize stock inventory, transportation between countries

An approach in 3 phases, consisting in exploring, testing and implementing

OBJECTIVES	PHASE 1: Collect and explore	PHASE 2: Test and assess feasibility	PHASE 3: Implement and industrialize
ELEMENTS OF METHODOLOGY	• Collect internal data sources • Get external data to complete them (product descriptions, market trends)	• Create a baseline model	• Test of the model: A/B test against baseline recommendations
	• Collaborate with business and marketing teams to **understand the objectives and expectations** for the recommendation system + analysis of the existing • Gather insights on KPIs	• Create **different models** and compare their performances • Choose the most performant model and improve it on an iterative way	• Implement CI/CD pipeline and API endpoints • Set up cloud-based or on-premise infrastructure • Integrate frontend • Implement KPIs for tracking recommendation effectiveness
	• Perform **exploratory analysis** (EDA) of the data	• Evaluate the performances of the models using **recall as metric**	• Conduct user acceptance testing • Automate retraining based on user feedback
PRACTICAL IMPLEMENTATION	• Data collection • Stakeholder interviews • Data cleaning and EDA • Initial Data Insights	• Develop and train models (collaborative filtering, ALS, deep learning models) • Compare performances using evaluation metrics • Optimize and validate	• **Deploy a cloud-based API** • Set up monitoring and logging • Automate CI/CD deployment
DELIVERABLES	• Data inventory report • Stakeholder insights report • Data summary presentation	• Baseline model and model comparison reports • Performance Analysis • Evaluation report	• Deployed **recommendation API** • **Monitoring Dashboard**

A mission over 3 months

1 Collect and explore

- Collect, clean and explore data
- Exchange with stakeholders

2 Test and assess feasibility

- Create some models and compare them
- Improve them on an iterative way

3 Implement and industrialize

- Test
- Implement pipeline
- Deploy frontend and get user-feedback

Kick-off – *Weekly meeting* – *Weekly meeting* – *Test of a 1st model* – *Weekly meeting* – *Weekly meeting* – *Weekly meeting* – *Final model chosen* – *Weekly meeting* – *Weekly meeting* – *Weekly meeting* – *Model deployed and industrialized*

Transforming Raw Data into Actionable Insights

Data preparation | Model training | Evaluation

Dataset Overview

- **1.12 million transactions**
- **Customer:**
  - Loyal customers: 66%
  - Prospect:18%
  - Inactive_1y: 15%
  - Top: 0.5%
- **Number of products in each transaction:**
  - Mean: 1.93
  - 50%: 2.00
- **Top 3 purchased categories (Quantity):**
  - Football 583k
  - Tennis 312k
  - Rugby 198k
- **Stock quantity:**
  - Mean:6.99
  - 50%: 3.00

Pre-processing steps

- **Remove** columns **Age** (70% missing data) & **Gender** ('C','N','U)
- Merge datasets: **'transaction' & ClientID & ProductID** (left join)
- Merge datasets: **stocks & products**
- Create **'CategoryFamily'** that combine 'Category' and 'FamilyLevel1'
  (e.g., 'Cycling-Helmet')
- **Customer segmentation** (Loyal & Top together, Inactive alone)
- **Train-Validation split**

1. Group transactions by ClientID
2. Split each client's purchase history:

   - 70% training (past purchases)

   - 30% validation (recent purchases)

Collaborative Filtering and LSTM models to recommend products

Data preparation → Model training → Evaluation

Collaborative Filtering (CF)
& Content-based filtering (CBF)

1. Embedding of the category + families of the products
2. Collaborative filtering to suggest 3 products that has weighted similarity to their purchased products
3. Look at which products are bought together (market basket analysis) and recommend 3 products based on the category
4. Recommend 1 product totally different from the ones bought by the customer, to detect if the customer has potential interest on unseen product
5. Total 7 products
   = 3 similar products + 3 paired products + 1 product

LSTM

1. Recognize patterns by grouping customer's past purchases as sequence, in chronological order
2. Encoding categories for model training
3. Process sequences to LSTM (Long-Short Term Memory) network and learn recurring shopping patterns
4. Train the model with CrossEntropyLoss & Adam optimizer
5. Output the top 3 most likely product categories for each user

Assessing Performance: Recall as the chosen metric

Data preparation → Model training → Evaluation

✅ Model Performance

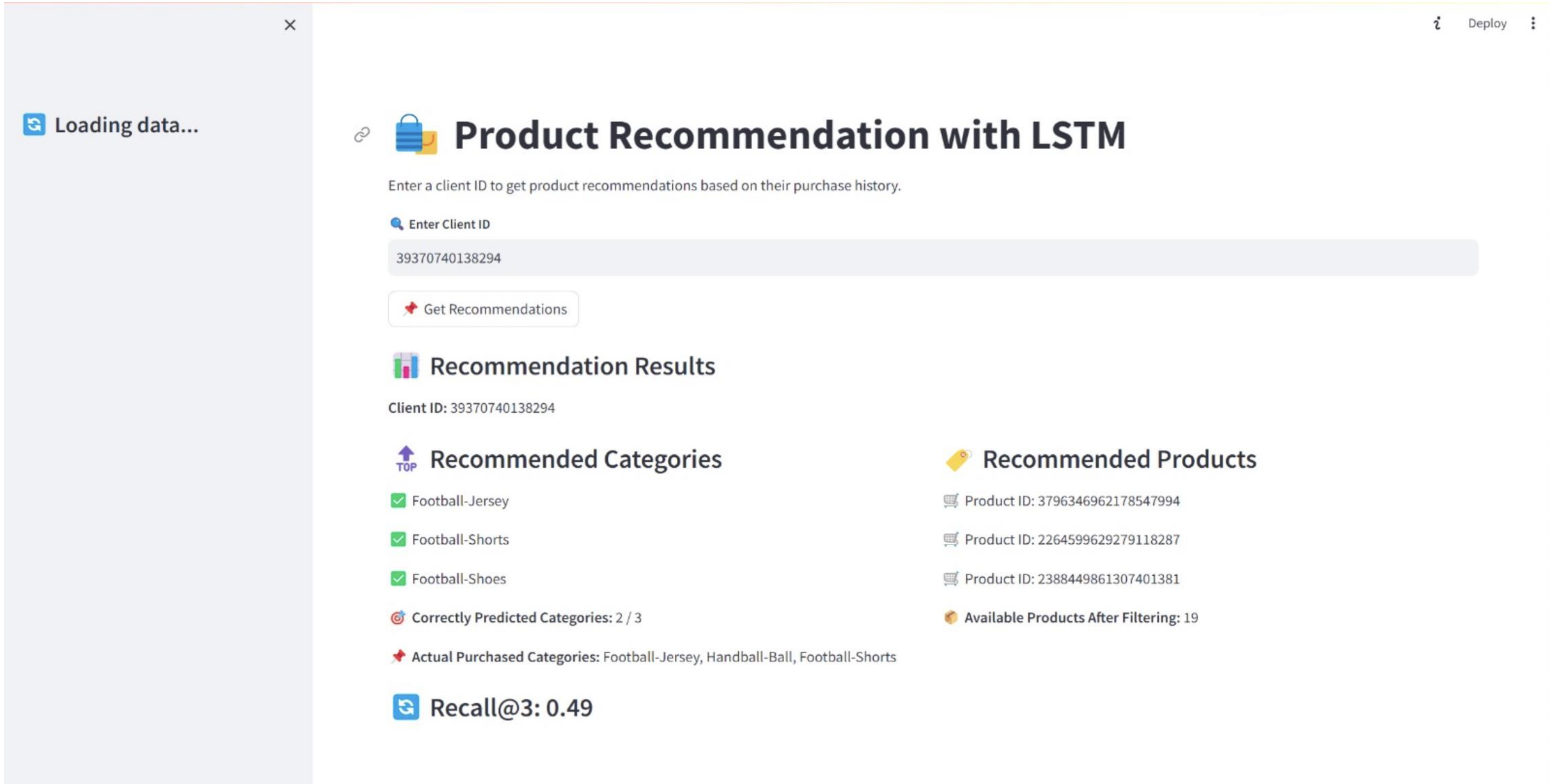
Recall@3: **73%** for 20 clients, **49%** for LSTM
Loss: 2.6 (value after model convergence)

✅ Key Insights

- Recommended products match customers' purchasing trends well.
- Positive impact on inventory optimization and cross-selling.

$$Recall@k = \frac{nb\ of\ products\ bought\ among\ the\ k\ recommendations}{nb\ of\ products\ bought\ by\ the\ customer}$$

Interactive Data Exploration: Visualizing with Streamlit



Some challenges due to the size of the dataset and the choice of the evaluation method

DATA

- **No rating** on the purchase made by the customers and no information about the products
- Few information about **product characteristics** (just name, category and family level)

COMPUTATIONAL CHALLENGES

- Try of **graph models** but crashed each time
- Try of applying **Sentence transformer**, then clustering models on the products and customers, but the models crashed. This is mainly due to the large size of the dataset

EVALUATION

- **Choice and computation of the metrics**: given the large number of products, slight variations in recommendation should not be considered as errors
- Choice of the method to **split the data** between training and testing sets (based on timeline / clients or just randomly)

Next steps

MODEL IMPROVEMENTS

Make our collaborative filtering model **scalable** (now only 20 clients)

Try to capture **more semantics** about the products, using **NLP** (like Sentence Transformer/Bert, ensuring that the models are deployable) and do some clusters of products

Search for more **efficient model** to optimize computation and reduce memory usage

Crawl the web to gather **characteristics** about the products

STRATEGY BASED ON RECOMMENDATION SYSTEM

Personalized email campaigns with tailored recommendations to present new products / regular personalized newsletters

Cross-selling: offer discounts if two products bought together (for products which are not well sold)

Events to promote the products and test them – invite users which would be interested by those products . E.g., sport merchandise for upcoming tournaments

Dynamic Pricing & Promotional Offers

STRATEGY FOR DIFFERENT CUSTOMER SEGMENTATIONS

PROSPECT

Collect data about the prospects if possible

- Online behaviour
- Survey to know more their preferences

If no information

- Recommend the most popular items
- Effort on advertisement

INACTIVE-USER

- **Send some personalized emails** to reignite interest
- **Personalized win-back discounts**

LOYAL USER

Early access to new product drops or exclusive collections

TOP USER

Curate personalized shopping experiences with premium selections

Team

Jinsuh YOU

MS DSBA CentraleSupélec & ESSEC
Background : Biomedical engineering

Binong HAN

MS DSBA CentraleSupélec & ESSEC
Background : Business

Julien DEHUT

MS DSBA CentraleSupélec & ESSEC
Background : Industrial engineering

Tonghan WEN

MS DSBA CentraleSupélec & ESSEC
Background : Business

Solène LEBOUCHER

MS DSBA - CentraleSupélec & ESSEC

Background: Business

Zongru SUN

MS DSBA CentraleSupélec & ESSEC
Background : Finance