Streamlining Machine Learning Lifecycle to Improve Retail Sales Performance Using MLflow

Sagar Baronia, Shubham Goyanka, Abhishek Jani, Aditya Roy Choudhary, Shirish Shinde, Matthew A. Lanham

Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

sbaronia@purdue.edu; sgoyanka@purdue.edu; jania@purdue.edu; roychoua@purdue.edu; shinde1@purdue.edu; lanhamm@purdue.edu

ABSTRACT

Organizations leveraging machine learning seek to streamline their machine learning development lifecycle. Machine learning model development possesses additional new challenges as compared to the traditional software development lifecycle. These include tracking experiments, code & data versioning, reproducing results, model management, and deployment. This paper aims to describe the implementation of MLflow on Azure databricks to streamline the process of building a recommender system to predict the user preference for a product or the likelihood of the user purchasing a product, given they are targeted with coupons in a promotional campaign. Finally, the entire machine learning pipeline is integrated with Flask using Rest API to serve the model on real-time and batch inferencing.

INTRODUCTION

Understanding factors that affect coupon redemption is crucial to determining the success of a promotional campaign. This paper discusses how our client can use millions of past transactions and demographics data to improve promotional campaigns efficacy using advanced recommender systems. However, getting a model into the real-world concerns more than building it. Deployment of the model into production is essential to take full advantage of the produced machine learning (ML) model; still only 22% of companies that use ML have successfully deployed an ML model into production. This paper discusses building and streamlining an entire ML pipeline and deploying the end model using a flask container using MLOps principles.

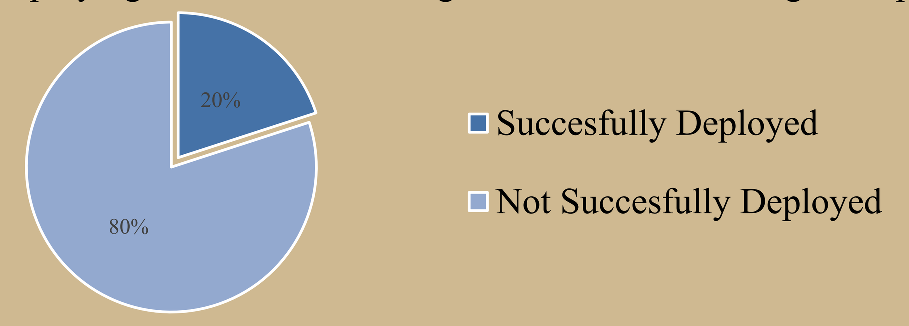

Fig 1. Proportion of Companies Successfully Deploying ML solutions

RESEARCH OBJECTIVES

- How can we leverage historical transactions and demographics data to build sophisticated machine learning system to improve promotional campaign efficacy?
- Is MLflow feasible MLOps tool to streamline the machine learning lifecycle?

LITERATURE REVIEW

Following resources were studied and referred to understand importance of coupon distribution optimization to improve retail performance, MLflow, MLOps, and Recommender Systems.

- Johnson et al., 2013
- Algorithmia, 2020
- Dr. Larysa et al., DDD Advisor
- Wang et al., 2020
- Doshi, S (2019)
- MLflow Documentation

Feature Name	ML Flow Support
Real-time & Batch Prediction	✓
GIT Integration	✓
Model Registry	✓
Role Based Access Controls	✓
Event Based Alerts & Notifications	✓
A/B testing and Multi-Arm Bandit Testing	×
Data Drift	×
Detection of Model Bias	×
Model Accuracy, Health and Prediction Stability	×

Table 1. ML Flow feature availability for deploying Recommendation Engine

METHODOLOGY

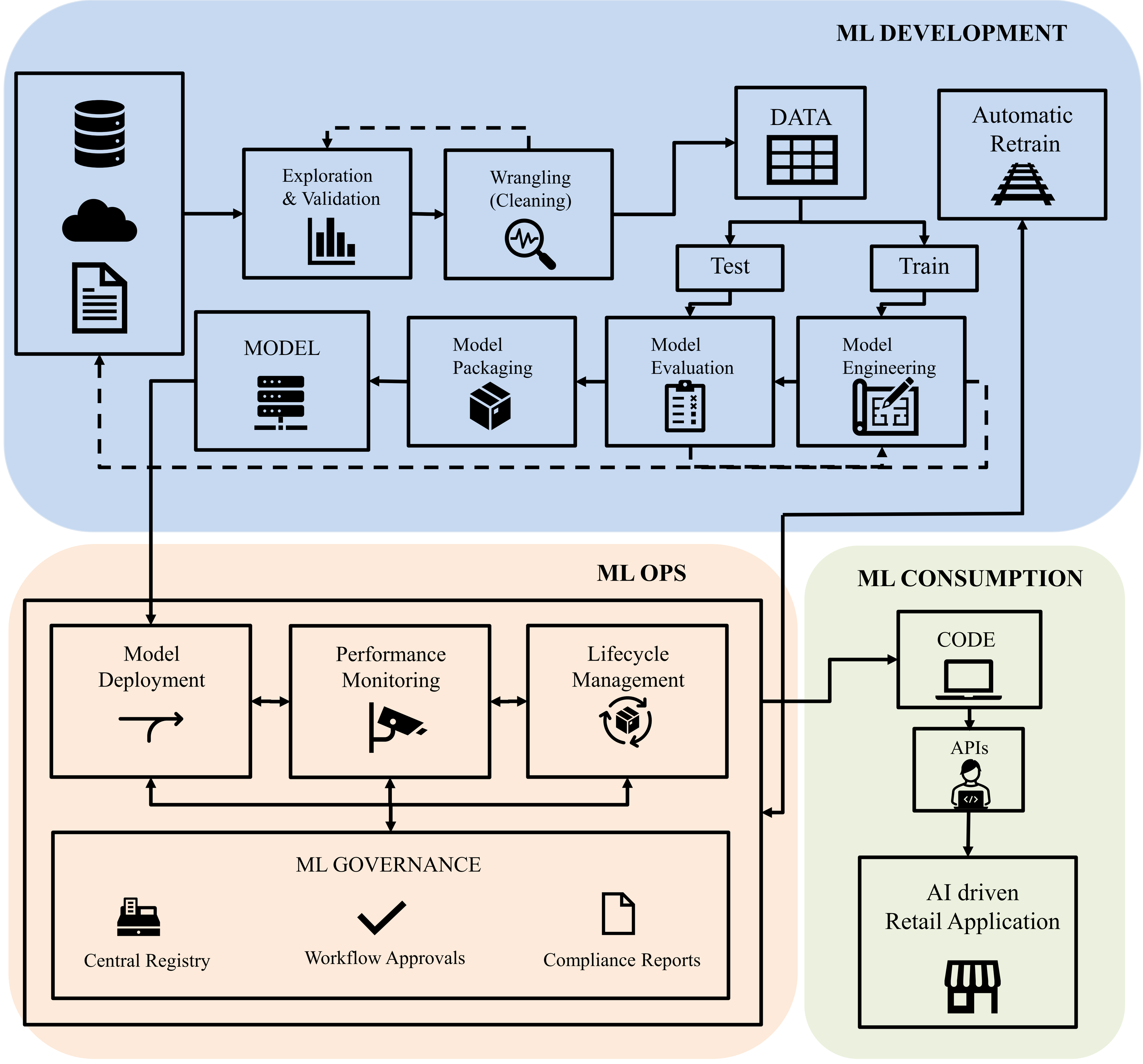

Fig 2. Architecture

STATISTICAL RESULTS

Extracting Top Products Recommendations for a Household using Collaborative Filtering Recommender System

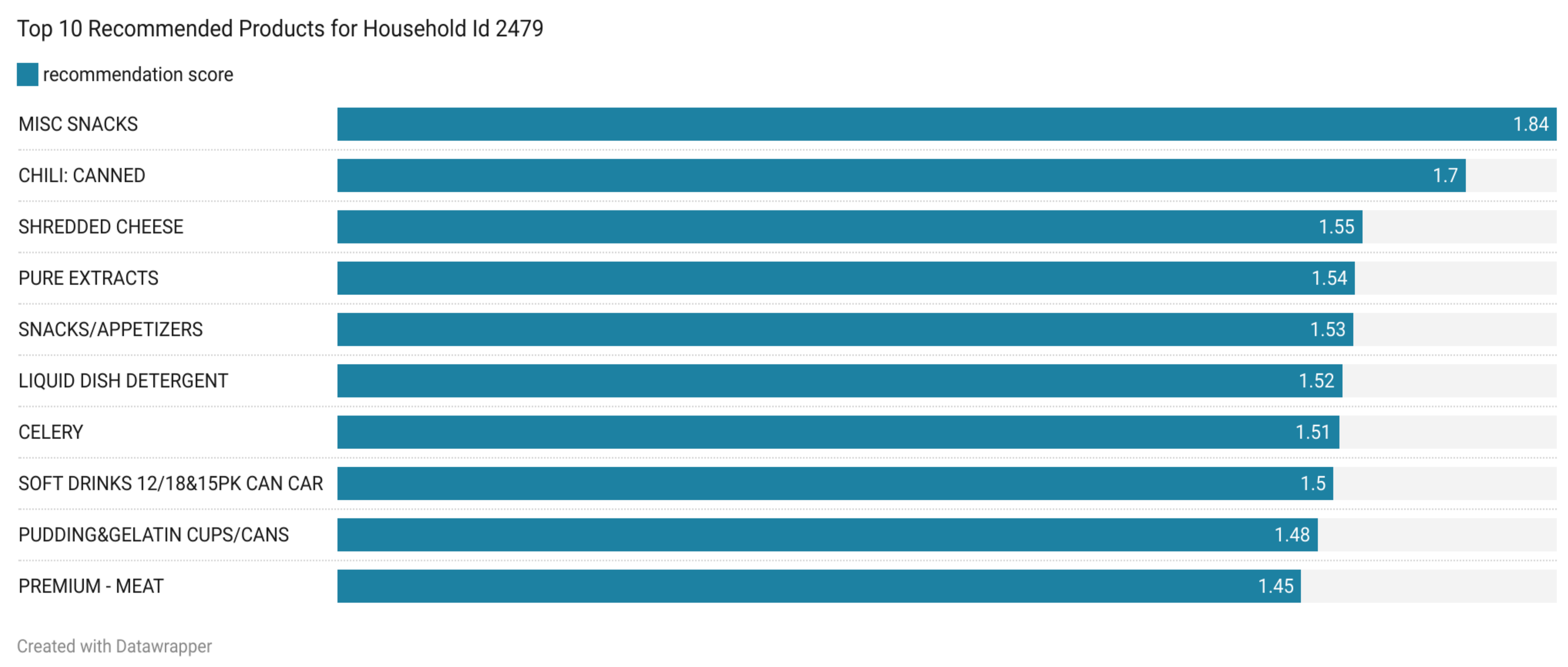

Fig 3.1. Model Results

Extracting Top Similar Products using Collaborative Filtering Recommendation System

Fig 3.2. Model Results

	Latent_Factors	Regularization_Rate	Iterations	Alpha	RMSE
0	40	0.2	100	15	1.616560
1	40	0.1	100	15	1.617299
2	40	0.3	100	15	1.617355
3	40	0.2	50	15	1.617561
4	40	0.1	50	15	1.617612
5	40	0.3	50	15	1.617821
6	40	0.1	100	25	1.621731
7	40	0.3	100	25	1.621938
8	40	0.1	50	25	1.622458
9	40	0.2	100	25	1.622668
10	40	0.2	50	25	1.622733
11	40	0.3	50	25	1.623312
12	20	0.1	100	15	1.625827
13	20	0.2	100	15	1.626155
14	20	0.2	50	15	1.626314
15	20	0.3	100	15	1.626318
16	20	0.1	50	15	1.626555
17	20	0.3	50	15	1.626704
18	20	0.1	100	25	1.633030
19	20	0.1	50	25	1.633109
20	20	0.2	100	25	1.633211
21	20	0.3	100	25	1.633325
22	20	0.2	50	25	1.633700
23	20	0.3	50	25	1.633758

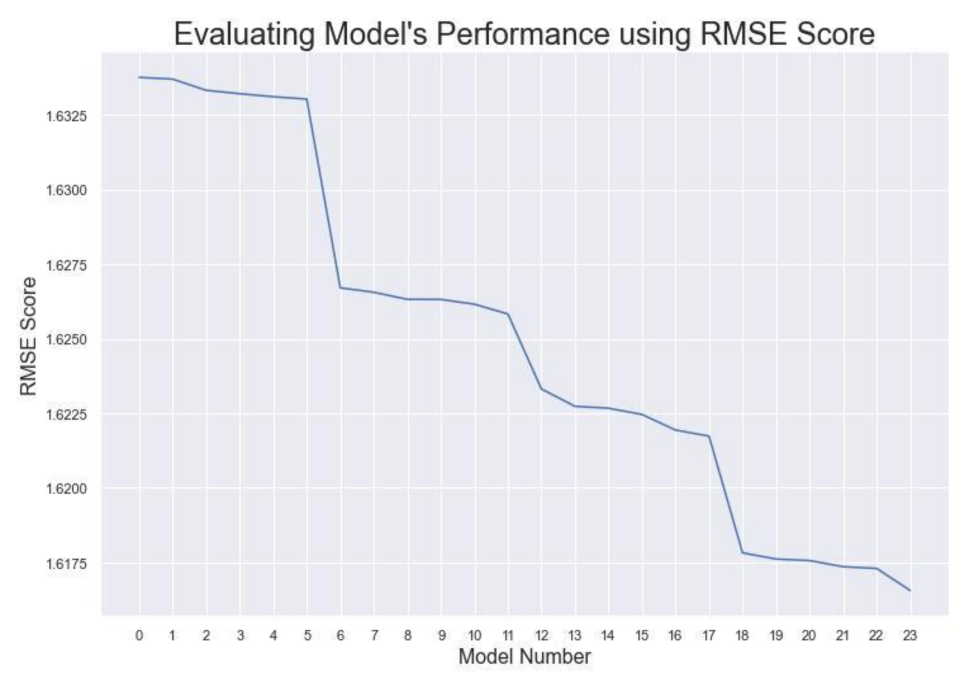

Fig 3.3. Model Results

BUSINESS IMPACT

Leveraging ML flow to implement recommender system on continuously expanding data reduced a lot of effort as ML Engineer against conventional method of CI/CD lifecycle resulting in saving work hours by more than 90% and establishing higher sense of reliability/robustness towards the process.

Using recommender system, we will generate the likeliness of a product to be purchased by a customer. Using that likeliness to broadly classifying into two categories in order to create an impact in business.

- ❖ Highly likely products to be purchased by a customer (0.9<p<1) based on probability will be shown to the customer with higher visibility to ensure our client secure business through these products
- ❖ Products that have likeliness between 0.7 to 0.9 to be purchased by a customer is the segment of products that our client will offer discount coupon

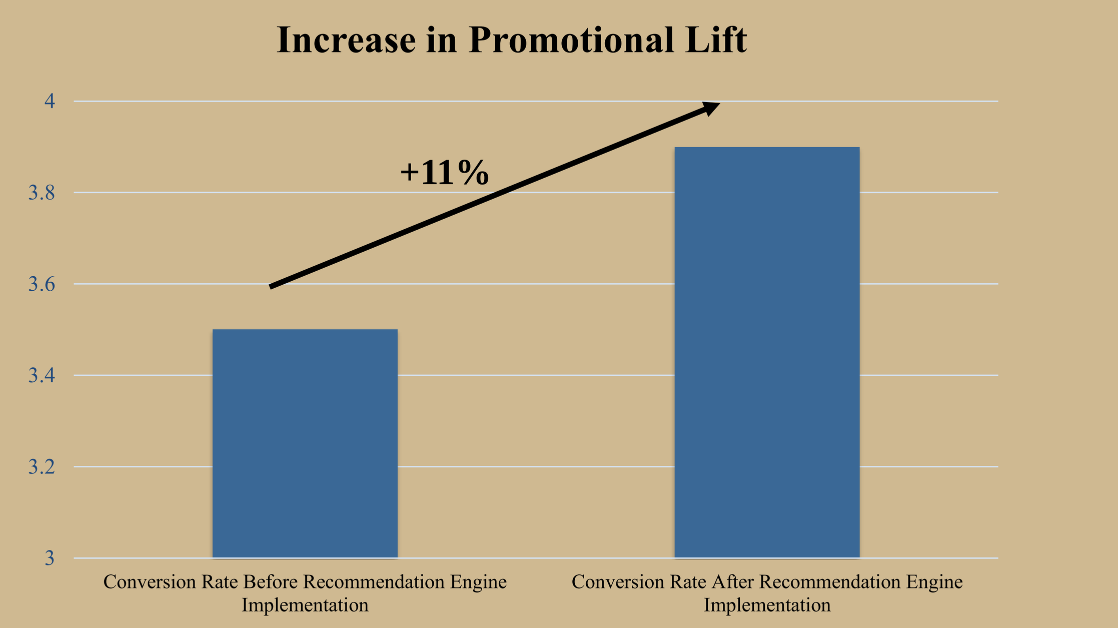

Fig 4. Promotional Lift

CONCLUSIONS

Streamlining the development of a recommender system using a hybrid collaborative filtering technique to improve retail sales performance using MLflow is feasible. Out of the many functionalities evaluated, role-based access controls, real-time prediction, batch inferences, and model registry are the top strength of MLflow.

Furthermore, many functionalities like Model Bias detection, Data drift, Concept Drift, and Resource Monitoring can be seamlessly integrated with ML flow using open-source libraries and tools like why logs, shap, Azure monitor, delta lake.

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham and our industry partner for this opportunity, their guidance, and support on this project.

PURDUE UNIVERSITY

Krannert School of Management

2022 informs BUSINESS ANALYTICS CONFERENCE HOUSTON