Modeling Business Cycle of Drugs

Data Science Institute
COLUMBIA UNIVERSITY

Zehui Wu, Hang Xu, Senqi Zhang, Shuyue Xu, Yajie Zhang
Industrial Mentor: Eric Matamoros Morales

Data Science Capstone Project with Novartis

Introduction

The sales volume forecasting for pharmaceutical companies is often tricky, in that drugs in different stages[1] often have varying sales performance . In this project, we seek to understand the effect business cycles[2] on drugs. Given information related to drugs and markets, we use machine learning models to identify the future market phases. Based on the predicted transition, we improve the performance of sales forecast.

Methods

To achieve the goal of predicting drug sales with cycles, we build a two-phase pipeline:

- In the first phase, we aim to predict the future stages for drugs given a specific time point. Drugs that have a growth-maturity transition before the specified time will be filtered and serve as the training data. Drugs that are in the stage of growth before the specified time point will be our target test set. We then fit the training data into a random forest model and generate transition points for all drugs in the test set.
- In the second phase, we aim to improve sales forecasting from the benchmark by damping the volumes after maturity. We train a regression model using a change rate for all the data points that are in the stage of maturity. Based on the transition points predicted, we generate the new forecasted volume using the benchmark.

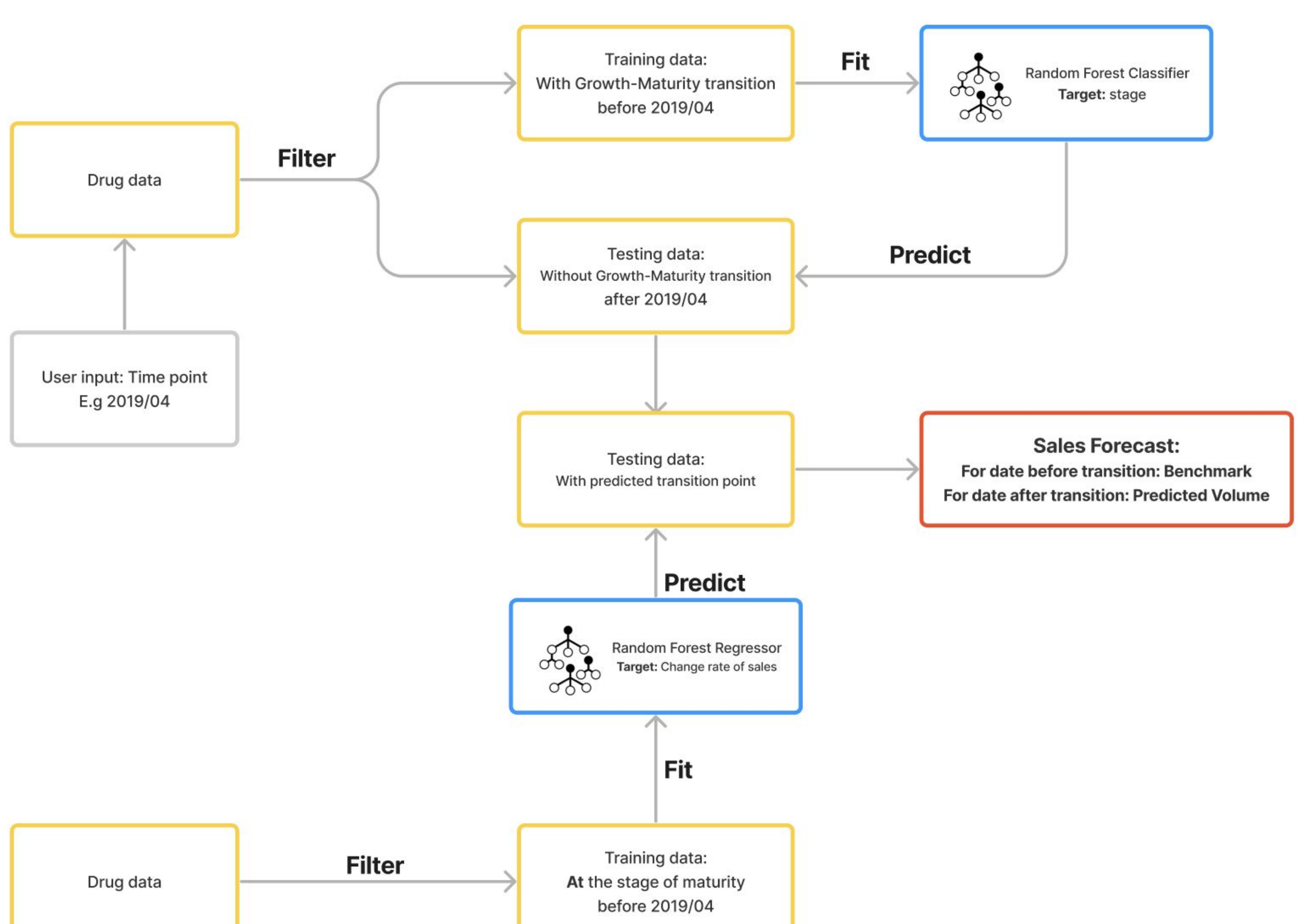

Figure 1. Pipeline

Results

We tried Random Forest, XGBoost, and DNN models for transition point estimation. It turns out that Random Forest performs slightly better than the others after we apply resampling on training data and change the prediction strategy. On average, the predicted transition is within 18 months of the actual transition point, which is better than the benchmark using the average length of a growth period.

After obtaining the predicted transition point, we applied two methods for volume prediction. The model that directly outputs volumes performs better than the model that outputs the difference between the benchmark and actual volumes. As a result, we achieved a 0.5% lift on MAE compared to the benchmark.

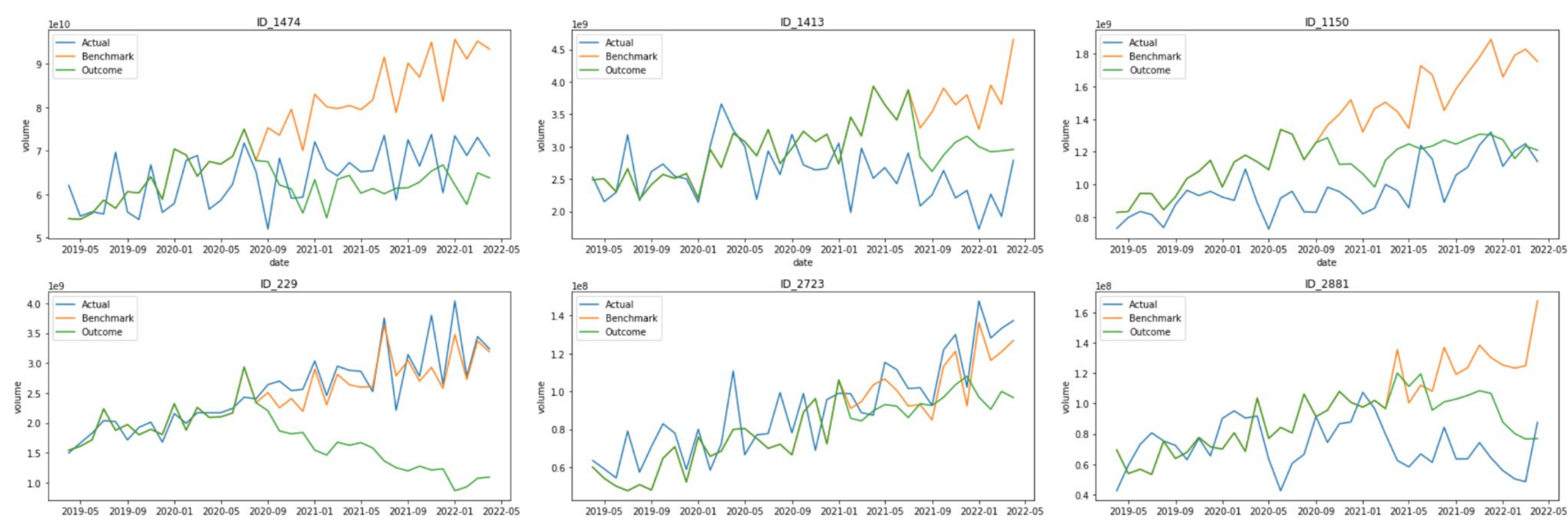

Figure 2. Sales Volume Forecast

Conclusion

We managed to get improvements from the the baseline methods by data sampling and hyperparameter tuning, although the given data is very noisy and most features do not have a strong prediction power. Still, the pipeline constructed can serve as guides for pharmaceutical companies' decision-making on marketing strategies and resource allocation. In addition to market phase identification, we also provide forecasted sales volume, which can be used to gain insights into future markets.

Acknowledgments

Thanks Novartis for providing the data and Eric for guiding the project.

Reference

1.Product Lifecycle in Pharmaceutical Industry. Article: https://www.happiestminds.com/wp-content/uploads/2020/12/Product-Lifecycle-in-Pharmaceutical-Industry-Journey-of-Drug-from-Ideation-o-commercialization.pdf

2. Connelly, F.J., Daignault, G. The life cycle concept as a long term forecasting model. JAMS 2, 455–464 (1974). https://doi.org/10.1007/BF02729389