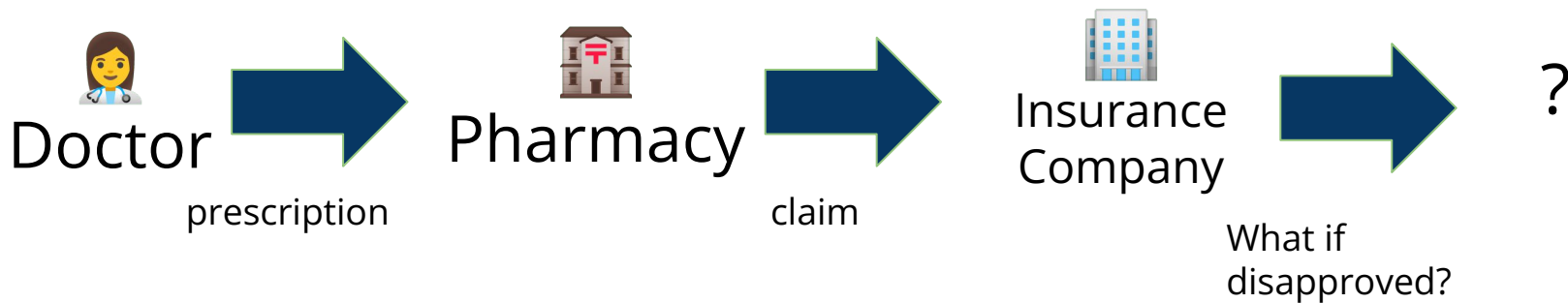

Team Pecan - CoverMyMeds

Po-Ying Chen
Lucas Mioranci
Pyongwon Suh

Project Description

Current prescription process is complicated



Goals

Based on the (simulated) pharmacy transaction data offered by CoverMyMeds,

1. Predict the patient's copay for a given prescription
2. Predict the formulary status of the prescription: will be rejected or not?

Expected Business Value

- Can expect how much they will pay at a pharmacy in advance(so that they will not come as a surprise)
- Prevent the situation that a claim is rejected and go back home with empty-handed
- Physicians can alter prescription in advance to reduce payments and avoid claim rejections
- Prevent the complicated process of filing prior-authorization, re-claim, etc.

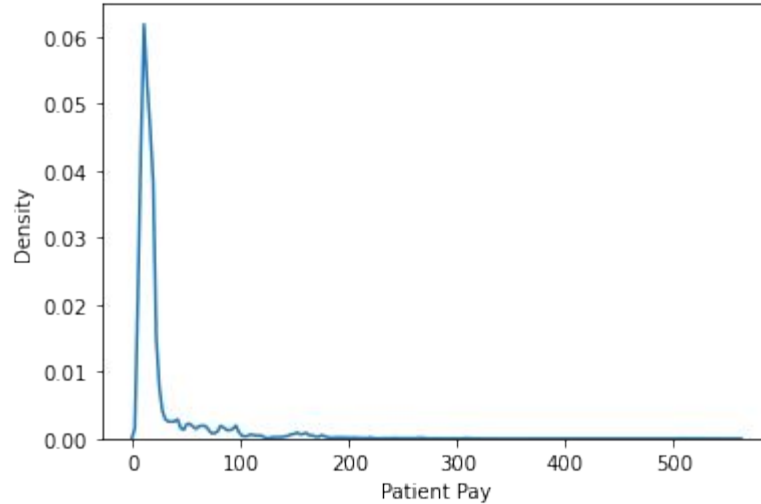
Data

~14 million entries of the following

	tx_date	pharmacy	diagnosis	drug	bin	pcn	group	rejected	patient_pay
0	2022-01-02	Pharmacy #6	G99.93	branded tanoclolol	725700	1UQC	NaN	False	13.39
1	2022-01-02	Pharmacy #42	U60.52	branded oxasoted	664344	NaN	52H8KH0F83K	False	7.02
2	2022-01-02	Pharmacy #37	Q85.91	branded cupitelol	725700	1UQC	NaN	False	13.39
3	2022-01-02	Pharmacy #30	U60.52	generic oxasoted	571569	KB38N	6BYJBW	False	10.84
4	2022-01-02	Pharmacy #18	N55.01	branded mamate	664344	NaN	ZX2QUWR	False	47.00

114 drugs, 133 diagnosis, 61 insurance plan types

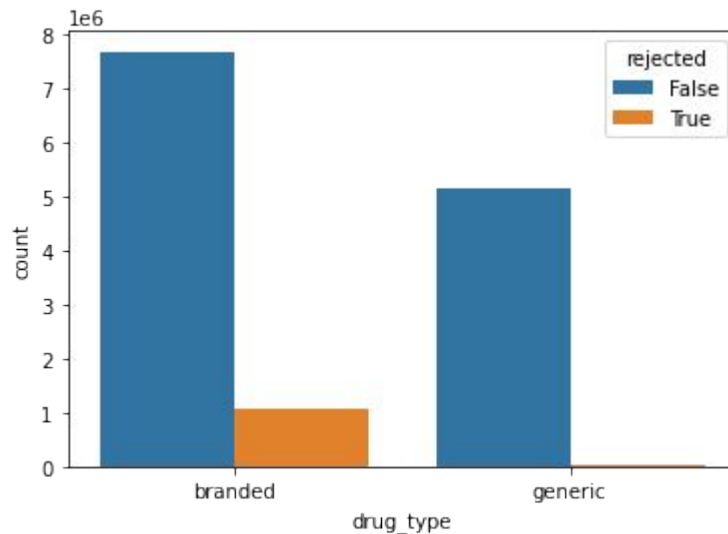
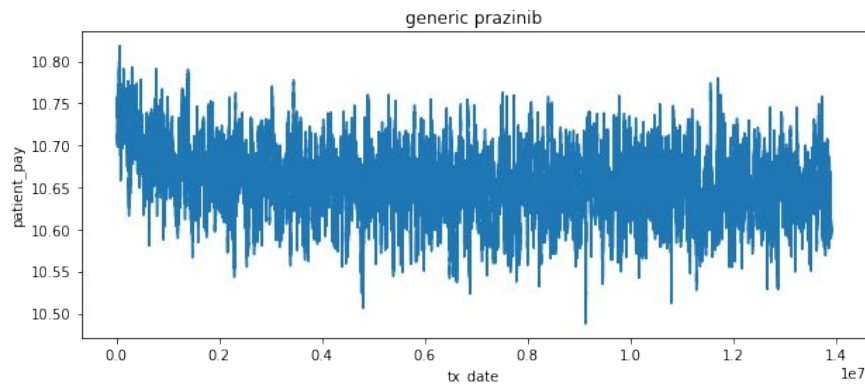
Payments distribution and rejection rates



Approximately 92.18% acceptance rate

Exploratory Data Analysis

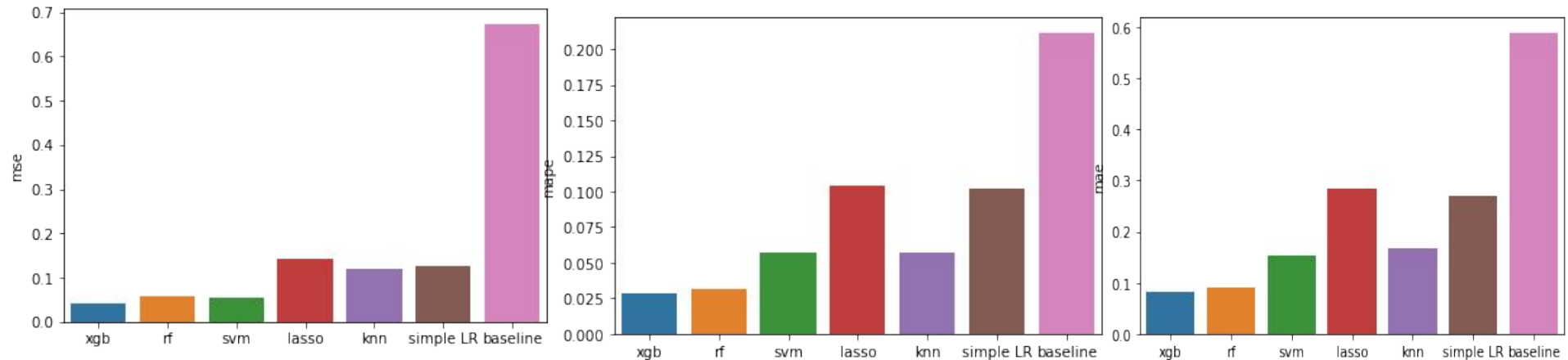
- Rejected claims: \$0 payments
- Drug type
- Insurance type(pcn), diagnosis



Prediction model for payments

- Tested various models: Simple Linear Regression, Regularized Linear Regression, KNN, Support Vector Machine, Random Forests, XGBoost
- Tested on sample taken from the whole data
- Used Mean Squared Error, Mean Absolute Percentage Error, Mean Absolute Error, R2 score as our metric
- GridSearchCV for hyperparameter tuning

Evaluations and Conclusion



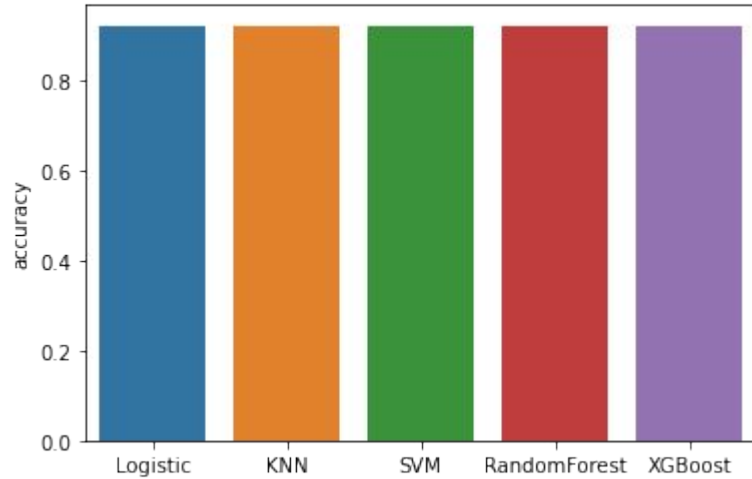
XGBoost performs the best

Prediction model for rejection

- Tested various models: Logistic Regression, classifier versions of KNN, Support Vector Machine, Random Forests, XGBoost
- Tested on sample taken from the whole data
- Used Accuracy as metric

Evaluation and Conclusion

Approximately 92.3% accuracy.



Future Directions

- Improvements on classification model needed
- Develop drugs clustering model physicians can use for altering prescription which are expected to be rejected or high in payments

Thank you!