Claims Denial Probability

Medical Insurance Claims Processor

- Analyzed the reasons for claims denial for the medical insurance claims over time
- Built a denial prediction model that estimates the probability of denial for each claim and highlights the top-3 potential reasons for denial, helping the team to focus on key areas and improve the overall efficiency of the process

Medical insurance claims processor needs to optimize the claims denial

Picture this…

You’re looking for a denial prediction model to identify the potential reasons for denial to reduce the time spent on processing medical claims manually. Currently, you are evaluating a potential solution to automate the claims process to increase the conversion rate (to approvals) and reduce the manual intervention, to enhance the efficiency of the workforce.

You turn to Accordion.

We partner with your team to analyze the reasons for claims denial for the medical insurance claims over time and to build a denial prediction model that estimates the probability of denial for each claim, including:

1) Estimating the probability of denial for the claims data using a binary classifier ensemble model built based on CatBoost algorithm, reducing the amount of manual effort in processing and identifying claims which would be denied
2) Identifying the top-3 reasons for denial for claims that are most likely to be denied using multiclass classifier ensemble model built based on CatBoost algorithm, narrowing down the corrective measures needed to be taken by the processor to convert the claim into an approval
3) Implementing the model into the client environment with an API endpoint to allow the end users to trigger the model when a new set of claims are received
4) Setting up processes to monitor the model performance, training the algorithm based on results to continuously learn to enhance the accuracy of the model and ensuring that the underlying patterns in the data are captured accurately

`

Your value is enhanced.

You have a model to flag denials or approvals with ~80% accuracy, reducing the time spent on processing the claims. You have a significant drop in the claims that were sent for processing as approvals with clerical errors, as the model is able to identify the errors and flag the appropriate reasons accurately. You also have the machine learning model that helps you identify rules in the data to drive denials.

CLAIMS DENIAL PROBABILITY

KEY RESULT

- ~80% accuracy reduced processing time

VALUE LEVERS PULLED

- Claims Denial Prediction Model
- Propensity Modeling

Reduced healthcare insurance denials for an insurance claims processor

Claims Denial Probability

Situation

- Client was evaluating a potential solution to automate the claims processing to increase the conversion rate (to approvals) and, also reducing the manual intervention during the process, thereby enhancing the efficiency of the workforce
- Partnered with the client to build a Denials Prediction model to identify the likelihood of denial for a medical claim and identify the potential reasons for denial to reduce the time spent on processing claims manually

Accordion Value Add

- Estimated the probability of denial for the claims data using a Binary classifier ensemble model built based on CatBoost algorithm, reducing the amount of manual effort in processing and identifying claims which would be denied
- Identified the Top-3 reasons for denial for claims that are most likely to be denied using Multiclass classifier ensemble model built based on CatBoost algorithm, narrowing down the corrective measures needed to be taken by the processor to convert the claim into an approval
- Implemented the model into the client environment with an API endpoint to allow the end users to trigger the model when a new set of claims are received
- Set up processes to monitor the model performance and train the algorithm based on results to continuously learn to enhance the accuracy of the model, and ensure that the underlying patterns in the data are captured accurately

Impact

- The model was able to flag denials or approvals with ~80% accuracy, reducing the time spent on processing the claims
- The client saw a significant drop in the claims that were sent for processing as approvals with clerical errors, as the model was able to identify the errors and flag the appropriate reasons accurately
- The machine learning model also helped the client identify rules in the data that drive denials

Approach and methodology – Features used

The features in the current data set can be broadly divided into four categories

Invoice Level Data

- Days between invoice open/create and service
- Invoice box-19 data availability
- HCPC code (*Healthcare Common Procedure Coding*)
- Quantity of equipment
- Invoice modifier code
- Item expected cost
- Item allowed cost
- Item charge cost
- Item tax cost

Policy Level Data

- Policy group (*Private/Government/ Commercial*)
- Policy company

Patient Data

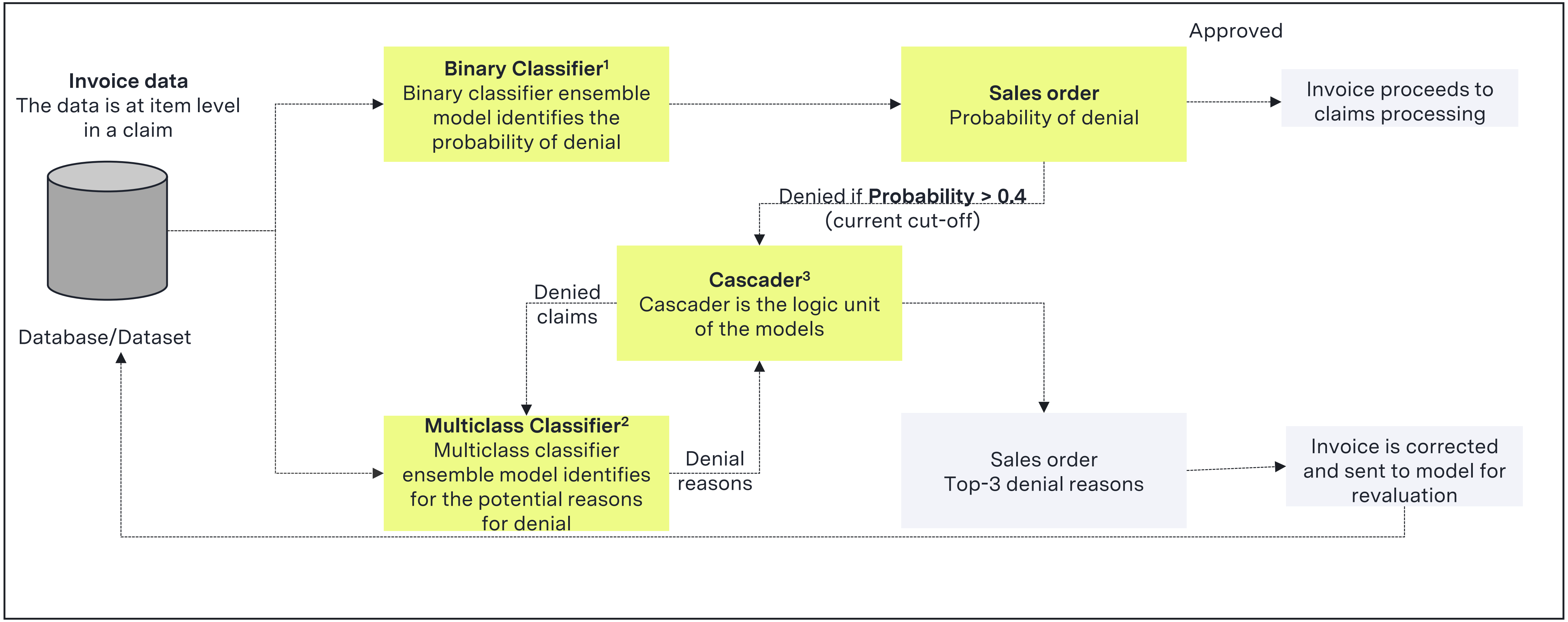
- Patient Age
- Patient City/State
- Patient Height
- Patient Weight
- BMI
- Patient ICD codes (*Disease codes*)

Doctor and CMN Data

- Availability of doctor details
- Doctor PECOS certification (*Provider Enrollment, Chain, and Ownership System*)
- Sleep therapy compliance status
- CMN (*Certificate of Medically Necessity*) details availability

Approach and methodology – Model flow

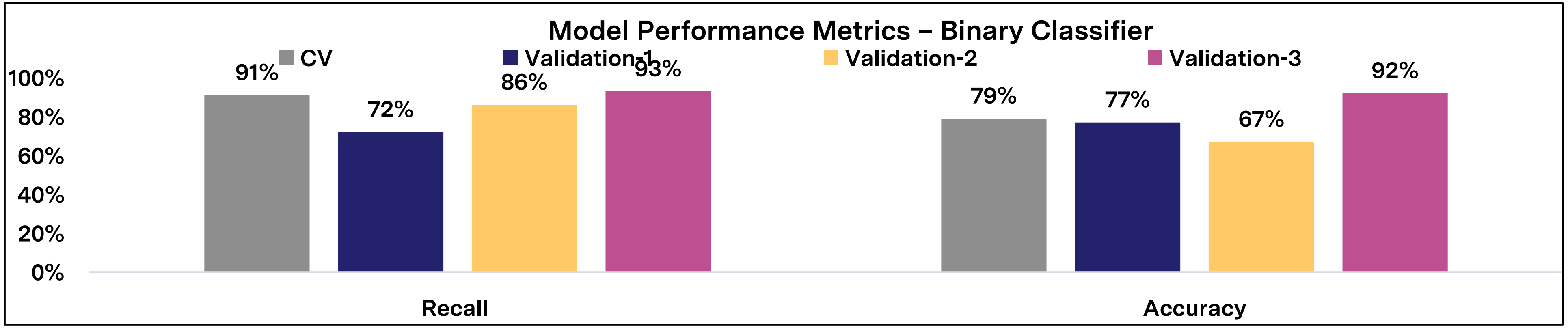
The model architecture consists of 3 parts. The Binary and Multiclass classifiers and the Cascader.

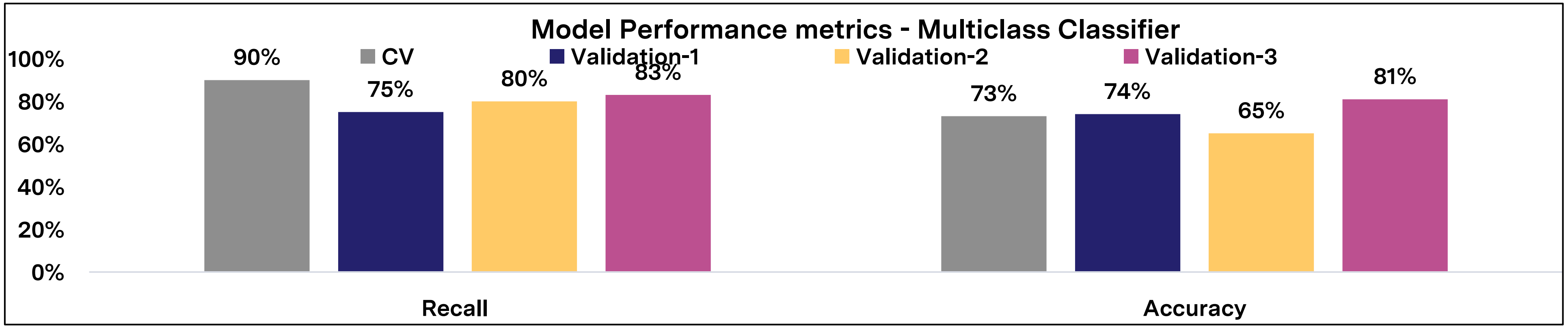


1 Binary Classifier: Built using CatBoost algorithm and optimized for binary predictions
2 Multiclass Classifier: Built using CatBoost algorithm and optimized for multiclass predictions
3 Cascader: Combines inputs from both models and provides the relevant outputs

Approach and methodology – Model results

The Binary Classifier captures denials with 91% Recall[1] on the CV[3] data and 86% on the Validation-2[4] data and with a corresponding Accuracy[2] of 79% and 67%. The Multiclass Classifier is on par in performance when compared to the Binary classifier.





1 Recall - % of actual denials that were predicted as a denial by the model
2 Accuracy - % of accurate predictions made by the model in the entire dataset
3 CV – original dataset used to train the model. The results shown are averaged out and are indicative of actual model performance
4 Validation - 1, Validation - 2, Validation - 3 – Different data sets used to test the model. Validation 3 is a biased dataset with 100% denials

