MATH490RH Project Report

Title: Optimizing Insurance Premiums: A Game-Theoretic Approach

Authors: Varun Gandhi, Adhiraj Chadha

Background

Insurance companies face the challenge of setting competitive premiums while minimizing potential losses. This problem is crucial in a highly regulated and competitive market, where balancing profitability with customer fairness is essential. Offering excessively high premiums may drive customers to competitors, while underpricing premiums increases the risk of financial loss for the insurance company. This project addresses this challenge using a combination of machine learning, game theory, and optimization principles.

The project’s main goal is to create a pricing strategy that allows an insurance company to set dynamic, individual-specific premiums. By leveraging a classification model and a game-theoretic decision framework, we aim to balance the trade-off between maximizing profit and staying competitive with a rival insurance company. This approach introduces a data-driven, adaptable methodology for pricing insurance products.

Methods

1. Risk Classification Using Decision Tree Classifier

The first step of the process involves risk classification. A Decision Tree Classifier was trained to predict the likelihood of an individual filing an insurance claim. This model was trained on a Kaggle insurance claims dataset containing customer attributes such as age, BMI, smoking status, and number of dependents. The target variable was binary, indicating whether the customer had filed a claim (1) or not (0).

Key Steps in Risk Classification:

- Data Processing: The dataset was cleaned and split into training and testing subsets to ensure model validity.

- Feature Selection: The key features used for training were age, sex, BMI, number of children, smoking status, and medical charges.

- Model Training: The Decision Tree Classifier was trained on the training subset and evaluated on the test set.

- Prediction: For each customer, the model predicted the probability of being classified as low-risk P(individual being low risk) or high-risk P(individual being low risk) based on their attributes. These probabilities served as inputs for the next step.

---

2. Game-Theoretic Framework Using Kuhn Game Tree

A diagram of a diagram

Description automatically generated

To model the decision-making process, we constructed an extensive-form game tree. The two players in this game are the insurance company and the "nature" (representing uncertainty in claim filing). The tree captures the possible actions (premium choices) and outcomes (accidents vs. no accidents) at each step. The terminal nodes of the tree contain the financial payoffs associated with each possible outcome.

Game Structure:

- Initial State: Insurance company predicts whether the specific individual is high risk or low risk based on the probabilities and a specific path is taken.

- Chance Nodes: Nature determines whether the individual files a claim or not based on whether they undergo an accident, with probabilities P(accident | predicted low risk) and P(accident | predicted high risk); and P(No accident | predicted low risk) and P(No accident | predicted high risk. We fix these probabilities based on the preexisting standards of the insurance companies.

- Payoffs: The terminal payoffs are the amounts the insurance company pays out in the event of a claim or receives as premium revenue if no claim is filed.

The expected payoff from the game tree is given by:

(-0.95\*premium + 500) \* P(low risk | attributes) + (2500 - 0.5\*premium)\*P(high risk | attributes)

This payoff function is used as the objective function for optimization.

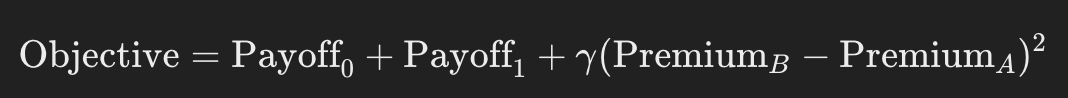
---

3. Optimization of Insurance Premiums

A graph of a curve

Description automatically generated

The objective function derived from the game tree serves as the basis for optimization. The goal is to determine the optimal premium (our company’s offered premium) that minimizes the expected payoff for each individual while staying competitive with a rival company’s premium The optimization problem is formulated as follows:



Explanation of Terms:

- P (low risk) and P(high risk) are the predicted probabilities from the Decision Tree Classifier.

- Premium\_A is the premium offered by the rival company.

- Premium\_B is the premium price we are trying to set.

- Gamma is a penalty parameter controlling the trade-off between profitability and competitiveness.

Optimization Process:

- Initialization: Start with an initial guess for our company’s offered premium price (typically 1000 for low-risk and 5000 for high-risk).

- Optimization Algorithm: We use the L-BFGS-B algorithm to minimize the objective function for each individual in the dataset.

- Dynamic Gamma: To increase adaptability, gamma is dynamically set for each individual based on their risk classification. For low-risk customers, gamma is higher to stay closer to the rival’s premium. For high-risk customers, gamma is lower to prioritize profitability.

Results

The optimized premiums varied based on customer attributes and risk classification. The use of dynamic gamma allowed for greater differentiation between high- and low-risk customers.

1. Risk-Based Premiums:

- High-risk individuals received higher premiums due to their increased likelihood of filing a claim.

- Low-risk individuals were charged lower premiums, and for some, the premium was close to the rival’s premium due to the competitiveness penalty.

2. Role of Dynamic Gamma:

- When gamma was high (for low-risk customers), our offered premium was closer to the rival’s premium to avoid losing the customer. We believe this individual is a low-risk individual and hence want to sell them the policy,

- When gamma was low (for high-risk customers), the model focused on setting a profit-driven premium. We believe this individual is high risk based on our classification and we are fine with charging them higher prices which is not competitive to the rival even if that means losing them to the rival.

In the image below 0.001 gamma assigned for high risk and 3 assigned to low-risk individuals dynamically.

A screenshot of a black screen

Description automatically generated

3. Key Insights:

- Dynamic Premiums: Premiums were adjusted for each customer, reflecting their individual risk and the rival’s premium.

- Trade-off Control: By tuning gamma the model effectively balanced profit maximization with competitiveness.

Conclusion

This project demonstrated a comprehensive approach to insurance premium optimization by combining classification, game theory, and optimization principles. The model produced dynamic, individualized premiums that balanced risk and competitiveness.

Summary:

- We classified risk using a Decision Tree Classifier.

- We modeled the payoff using a game-theoretic tree and derived an objective function.

- We optimized premiums to minimize losses while staying competitive with a rival’s premiums.

Future Work:

- Dynamic Rival Pricing: Model how the rival’s premium changes based on Company B’s pricing strategy.

- Improved Classifier: Use advanced machine learning models to better predict risk probabilities.

Acknowledgments:

We acknowledge the Kaggle insurance claims dataset for providing the data necessary to train the Decision Tree Classifier. We thank Professor Flaherty for guiding us and giving us the idea of using Game trees to solve this problem.

References

1. Kaggle. "Insurance Claim Prediction Top ML Models." Available at: <https://www.kaggle.com/code/yasserh/insurance-claim-prediction-top-ml-models/output>