# **Smart Underwriting for Home Inspections**

Ashvin Raj, Chu-Yun Hsiao, Raghav Chadha, Xiaoyu Guan, Yi-Chun Huang, Khan, Shoaib Amjad

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

raj53@purdue.edu; hsiao56@purdue.edu; chadha24@purdue.edu; guan115@purdue.edu; huan2081@purdue.edu; khan180@purdue.edu

## **ABSTRACT**

This project explores the integration of data analytics and machine learning in optimizing the home insurance inspection process, conducted in collaboration with a major insurance company. Guided by key questions surrounding the performance and improvement of the current inspection algorithm, the study leverages a mixed-methods research design, combining quantitative analysis of home inspection data with expert insights. This advancement is crucial as it addresses the need for more accurate and efficient risk assessment in the evolving insurance industry, where data analytics plays a pivotal role in shaping operational strategies. We employed a combination of predictive modeling techniques, harnessing through Python. These methods allowed for a more sophisticated analysis of risk factors, leading to a more targeted approach in identifying properties requiring inspection, thereby optimizing resources and enhancing customer satisfaction. The results demonstrated the superiority of the Random Forest model, which achieved the highest ROC AUC score of **0.89**. The metric indicates the model's robustness and reliability in accurately classifying both positive and negative instances, making it suitable for scenarios that demand high accuracy and the ability to correctly identify properties requiring inspections and effectively minimizing redundant underwriting expenses.

**Keywords**: Insurance, Home Inspection Process, Predictive Modeling, Data Analytics

## **INTRODUCTION**

The insurance industry is undergoing rapid transformation, driven by the integration of data analytics and machine learning. As highlighted by Gartner, WSJ, and Forbes, predictive analytics is reshaping risk assessment practices, driving customer-centric approaches, and streamlining operations across the insurance landscape. In the property insurance sector, the need for innovative data-driven solutions is critical.

This project, conducted in partnership with a major insurance company, seeks to optimize its home insurance inspection process. With evolving economic conditions and inflationary pressures, the company's current inspection algorithm requires enhancement to accurately reflect the changing landscape. The overarching goal is to develop an improved algorithm that effectively prioritizes home inspections, leading to better resource management and optimized underwriting decisions.

The project is guided by the following key questions:

1. How does the current inspection algorithm perform, and what areas require improvement?
2. What are the critical factors contributing to the need for home inspections, and how can these be leveraged to create a more accurate predictive model?
3. What proactive measures can be employed to identify high-risk properties?

Addressing these questions offers direct benefits to the insurance company. A more refined inspection process translates to improved risk management, accurate premium pricing, optimized resource allocation, and ultimately, increased customer satisfaction and profitability.

This project will analyze the current algorithm's performance metrics with the aim of identifying areas for improvement. Subsequently, a new machine learning model will be developed, incorporating key data elements to enhance the accuracy of home inspection predictions. The focus will be on pinpointing features that strongly correlate with inspection outcomes and high-risk properties.

We anticipate this project will demonstrate the transformative power of data analytics and machine learning in the insurance industry. By streamlining the home inspection process, the insurance company can reduce operational costs, refine risk assessment, and deliver a superior customer experience.

## **LITERATURE REVIEW**

|  |  |  |
| --- | --- | --- |
| **Study** | **Motivation** | **Algorithm** |
| Jae Joon Ahn et al., 2012 | Enhance real estate appraisal forecasting | Ridge Regression |
| Ruth M. Ogunnaike et al., 2017 | Prediction of insurance claim | Linear Regression,  Random Forest,  Support Vector Regression,  Feed Forward Neural Network |
| Abhishek Chauhan et al., 2020 | Evaluation of real estate appraisal | Bootstrap Aggregating,  Random Forest,  Adaptive Boosting,  Gradient Boosting |
| Jacqueline Schmitt et al., 2020 | Predictive model-based quality inspection | Artificial Neural Networks,  Support Vector Machines,  Decision Trees |
| Taha et al. 2022 | Examine the importance of feature selection for improve model performance and influential features for business insights | Neural Network,  Stochastic Model |
| Cedric Seger, 2018 | Compare the performance between one hot encoding, binary encoding and feature hashing | Feed Forward Neural Network,  Logistic Regression |
| Kumar, 2013 | Prediction of Consumer Purchase Decision | Logistic Regression |
| Ramraj et al., 2017 | Experimenting XGBoost Algorithm for Prediction and Classifi cation of Different Datasets | Gradient Boosting,  XGBoost |
| **Our Study** | **Feature identification and information gain for optimizing underwriting process** | **Logistic Regression,**  **XGBoost,** **Feature Importance** |

Table 1: Literature Review

Real estate appraisal and prediction have been the subject of extensive research, with scholars exploring various methodologies and technologies to enhance accuracy and efficiency. This literature review synthesizes findings from relevant studies, each addressing distinct aspects of real estate valuation and prediction.

Jae Joon Ahn et al. introduces the GA-Ridge method for real estate appraisal forecasting, integrating ridge regression with genetic algorithms to address predictor variable selection challenges and improve upon traditional regression techniques. Experimental validation on the Korean real estate market demonstrates the effectiveness of GA-Ridge in forecasting, offering insights into enhancing ridge regression for practical applications.

Similarly, Ruth M. Ogunnaike et al. focuses on predicting severity loss values in insurance claims through machine learning regression techniques. Leveraging high-dimensional data from Allstate insurance company, the study compares the performance of Linear Regression, Random Forest Regression, Support Vector Regression, and Feed Forward Neural Network models. The findings highlight the efficacy of ensemble methods, particularly bagged Random Forest Regression, in achieving accurate predictions, emphasizing the relevance of machine learning in insurance risk assessment.

Abhishek Chauhan et al. investigates the accuracy of real estate price prediction in Bengaluru through ensemble machine learning methods. Leveraging techniques such as Bootstrap Aggregating, Random Forest, Adaptive Boosting, and Gradient Boosting, the study showcases the effectiveness of ensemble methods, particularly Gradient Boosting, in improving prediction accuracy and reducing error rates, suggesting their potential for enhancing real estate appraisal methodologies.

Jacqueline Schmitt et al. explores the integration of machine learning and edge cloud computing for quality inspection in industrial manufacturing. Conducted by Schmitt et al. (2020), the study presents a holistic approach to quality inspection, integrating data acquisition, modeling, and technological implementation. Through a case study in Surface Mount Technology (SMT) manufacturing, the research demonstrates significant improvements in inspection efficiency and economic benefits, highlighting the potential of machine learning and edge computing in optimizing manufacturing processes.

Furthermore, Taha et al. emphasizes the importance of feature selection in enhancing the accuracy and efficiency of machine learning models within the insurance industry. Highlighting the complexity of actuarial science and the potential of advanced ML techniques such as neural networks, the study underscores the significance of feature selection in refining models for more accurate predictions.

Cedric Seger examines the feasibility of binary encoding as an alternative to one-hot encoding for categorical variable encoding. While one-hot encoding offers advantages, such as applicability in models requiring numerical input, binary encoding presents lower dimensional representations but may exhibit lower performance, especially on certain datasets.

Kumar investigates the fitness of logistic regression in analyzing categorical variables and predicting outcomes related to consumer purchase decisions. Leveraging demographic variables, logistic regression offers insights into consumer behavior, crucial for informing marketing strategies effectively.

Lastly, Ramraj et al. aims to quantitatively compare the accuracy and speed of the XGBoost algorithm with Gradient Boosting using various datasets. By applying both models to the same datasets, the study seeks to evaluate their accuracy and Mean Squared Error (MSE) performance, offering insights into their relative effectiveness and efficiency.

In conclusion, these studies collectively demonstrate the diverse applications of advanced methodologies, such as machine learning, neural networks, and ensemble methods, in real estate appraisal, insurance risk assessment, and quality inspection in manufacturing. The findings underscore the transformative impact of technology-driven approaches on accuracy, efficiency, and decision-making across various domains, paving the way for future advancements in predictive analytics and decision support systems.

## **DATA**

Our analysis aimed to extract valuable insights from various data sources provided by the insurance client to support strategic decision-making processes. We meticulously sifted through and integrated multiple data tables, each containing critical information pertinent to our study. The primary data sources utilized in our analysis were:

* **CoverageA\_SCD2:** A slowly changing dimensions (SCD) table tracking historical changes in policy coverage details. This table captured snapshots of coverage information over time, enabling us to analyze policy coverage evolution and identify potential trends or anomalies.
* **Dwelling\_SCD2:** Another SCD table, designed to maintain a comprehensive record of changes in residential property attributes. By monitoring alterations to dwelling characteristics, such as renovations or additions, we could assess their potential impact on risk profiles and pricing strategies.
* **Policy:** A centralized repository containing core policy-related information, including policy numbers, effective dates, and associated customer details. This table served as a crucial link for integrating data from various sources.
* **Quote Stage Home Inspection Data:** Acting as our base table, this dataset comprised detailed inspection data collected by sales agents during the initial quotation stage. These records provided valuable insights into the condition of properties, enabling us to evaluate potential risks and tailor our analysis accordingly.

**Linking Data Tables:** To effectively leverage the information contained within these disparate data sources, we developed a robust methodology to link the tables using unique policy numbers as the common key. This integration process was crucial for combining and comparing data across different sources, enabling a comprehensive view of each policy's lifecycle and associated risk factors.

**Data Preprocessing:** During the data preparation phase, we encountered several challenges, including missing values in critical columns. We conducted thorough investigations to understand the underlying reasons for these data gaps and employed appropriate imputation techniques to fill in missing values with statistically sound estimates. This ensured the integrity and accuracy of our analysis, minimizing the impact of incomplete or inconsistent data.

A significant hurdle was keeping track of updates in the inspection data and the Slowly Changing Dimensions (SCD) tables. These tables captured historical changes, necessitating careful attention to policy numbers and the corresponding dates when changes were recorded. We implemented robust data management processes to ensure no important updates were overlooked, maintaining a comprehensive view of each policy's evolution over time.

In summary, our approach was methodical and thorough, combining meticulous data integration, quality assurance, and temporal analysis. By identifying and leveraging key data sources, addressing data quality issues, and paying close attention to changes over time, we were able to effectively link related data and extract valuable insights. This systematic process not only addressed the specific analytical needs of our project but also demonstrated the broader value of detailed and rigorous data analysis in enhancing decision-making processes within the insurance industry.

## **METHODOLOGY**

The study employs a mixed-methods research design, combining quantitative analysis of home inspection data collected at quote stage with slowly changing data insights from home inspection. A comprehensive dataset of home inspection reports is obtained from sales agents. The dataset includes information on various aspects of household conditions, such as structural integrity, fire protection, pets, roofing, and overall property dimensions and condition.

By conducting Exploratory Data Analysis (EDA), correlations and interactions between variables were explored to uncover relationships and inform feature engineering strategies. Based on insights from EDA and expert consultations, we created new features derived from existing variables or external data sources to enhance the predictive power of the model.

In the data preprocessing phase, missing values were addressed through imputation or deletion, while outliers in numerical features like Coverage A amount were treated using statistical methods. Continuous variables, including Coverage A amount, were discretized into bins to simplify interpretation. Categorical variables were encoded numerically, employing techniques like one-hot encoding or label encoding. Additionally, numerical features were normalized to ensure uniform scales across variables, preventing any single feature from dominating model training. These preprocessing steps collectively enhanced the dataset's quality, standardized its format, and prepared it for effective model training and evaluation.

For classifying the binary target variable, indicating whether a property is high risk or not, we will apply random forest, XGBoost and logistic regression models. These models are chosen for their ability to handle binary classification tasks effectively and efficiently.

Those models will be trained using the labeled training dataset, and their performance will be evaluated using appropriate evaluation metrics such as precision, recall, F1-score, and AUC-ROC (Area Under the Receiver Operating Characteristic Curve).

The model development process involves the following steps:

**Feature Selection**: Identify relevant features and preprocess the data to prepare it for model training.

**Model Training**: Fit the random forest, XGBoost and logistic regression models to the training data, tuning hyperparameters as needed to optimize performance.

**Model Evaluation**: In the context of imbalanced datasets and binary classification tasks, a 70/30 split ratio allows for a sufficient amount of data to train the models while still reserving a sizable portion for independent evaluation. Stratified K-Fold Cross-Validation was employed during model training to ensure that each fold's class distribution is representative of the overall dataset. It helps mitigate the risk of having highly imbalanced class distributions in individual folds and further assess performance and reduce the risk of overfitting.

**Comparison**: Compare the performance of the logistic regression, random forest and XGBoost models based on evaluation metrics to determine the most effective classifier for predicting property risk.

**Model Interpretation**: Interpret the coefficients and feature importance scores from the logistic regression, random forest and XGBoost models to gain insights into the factors influencing property risk classification.

By applying logistic regression, random forest and XGBoost models, we aim to develop robust classifiers that accurately predict whether a property is high risk or not, based on data-driven insights derived from the home inspection dataset and underwriting information.

Potential limitations of the study include sample biases in the home inspection and underwriting datasets, the inherent subjectivity of home inspection assessments, and the complexity of underwriting decisions influenced by multiple factors beyond home condition. These limitations are acknowledged and discussed in the interpretation of findings and recommendations for future research.

## **MODEL(s)**

In our research, we carried out feature engineering to have effective variables which could be used in the models. Initially, we calculated the duration of employment for each agent by determining the years since their hire, derived by subtracting the hire year from the current year (2024). This computation resulted in a new numerical feature, 'YearsSinceHire', which quantifies the tenure of each agent. We analyzed the distribution of this data to gain insights into the tenure diversity among the agents.

Furthermore, to understand the variations in coverage amounts, we implemented a custom binning strategy on the 'CoverageAAmount' feature. We defined specific range bins to categorize the coverage amounts into intervals such as '0-100k', '100k-125k', etc., and labeled them accordingly. This categorization helped in examining the distribution and frequency of different coverage amounts, enabling a structured analysis of how coverage is allocated among different segments.

We also had other custom variables made such as 'Deductible\_Binned', 'ReplacementCost\_Binned', 'ConstructionYear\_Era'. Thes features were critical in facilitating a more detailed examination of the dataset, aiding in the discovery of underlying patterns and trends that could be significant in the broader context of our research.

The model building process began with defining a set of predictor variables that included both categorical and numerical features such as 'ConstructionType', 'PolicyTypeID', and 'YearsSinceHire'. The target variable for our predictive model was 'Target\_Changes'.

We utilized the train-test split method to divide the dataset, with 80% allocated for training and 20% for testing, ensuring a randomized and representative division through the specification of a seed (random\_state = 42). The preprocessing pipeline employed two main transformers: StandardScaler for normalizing numerical variables and OneHotEncoder for encoding categorical variables. This setup helps in handling the diverse data types effectively.

For modeling, we tried an XGBoost classifier, configured with specific parameters like learning rate and maximum tree depth, optimized through prior experiments. The model was encapsulated within a pipeline that sequentially applied preprocessing and classification, ensuring a clean workflow and avoiding data leakage.

We also implemented a RandomForestClassifier to predict 'Target\_Changes' based on an extensive set of predictor variables. e preprocessing framework was meticulously organized to effectively manage diverse data types: StandardScaler normalized continuous variables, OneHotEncoder was applied to categorical variables requiring dummy coding, and OrdinalEncoder was utilized for categorical variables where ordinal relationships exist.

The choice of RandomForestClassifier was driven by its robustness in preventing overfitting and its capacity to manage both linear and nonlinear relationships, which is critical for handling complex datasets like ours. RandomForest operates by constructing multiple decision trees during training and outputs the class that is the mode of the classes predicted by individual trees, hence improving the overall accuracy and stability of the model.

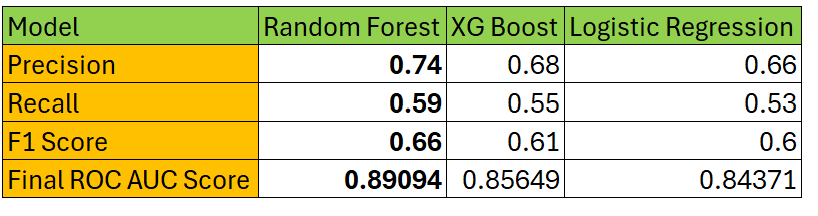
For our RandomForest model, we utilized 100 trees (n\_estimators=100), ensuring sufficient model complexity and learning capability without excessively increasing computational load. The 'n\_jobs=-1' parameter was used to enable parallel processing, significantly speeding up the training process by utilizing all CPU cores. The 'random\_state=42' ensured that our results were reproducible, with the same split of data and the same sequence of operations applied each time the code is executed.

We also tried the logistic regression model. The Logistic Regression was selected for its simplicity and effectiveness in binary classification problems. It operates by estimating probabilities using a logistic function, which is particularly useful for cases where you need to provide a probabilistic framework and understand the influence of individual predictors. This model is intrinsically interpretable, making it easier to explain the results and the influence of each feature on the outcome.

For our Logistic Regression model, key parameters were carefully chosen to optimize performance. The regularization strength (C parameter) was set to prevent overfitting, with lower values increasing the regularization effect to avoid overly complex models that do not generalize well. The solver used in our model was 'liblinear', which is well-suited for small to medium datasets and binary classification problems.

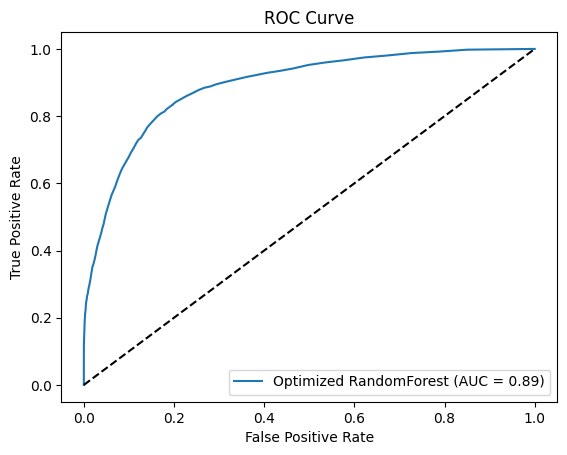
## **RESULTS**

We evaluated the model's performance with multiple metrics: confusion matrix, ROC-AUC score, precision, recall, and F1 score. These metrics collectively provide a comprehensive evaluation of the model's performance, emphasizing its accuracy in class prediction and its ability to balance sensitivity and specificity. The table below presents a summary of the performance metrics for each model:



The Random Forest model demonstrated superior performance across all metrics, achieving a precision of 0.74, recall of 0.59, an F1 Score of 0.66, and the highest ROC AUC score of 0.89094. This indicates its robustness and reliability, making it suitable for scenarios that demand high accuracy and the ability to correctly identify both positive and negative classes.

XG Boost, while slightly trailing behind, showed commendable performance with a precision of 0.68, recall of 0.55, and an F1 Score of 0.61, accompanied by a ROC AUC score of 0.85649. Its results suggest it is a strong candidate for situations where the balance between detecting positives and minimizing false positives is critical. Meanwhile, Logistic Regression offered competitive yet slightly lower metrics with a precision of 0.66, recall of 0.53, an F1 Score of 0.60, and a ROC AUC score of 0.84371. The simplicity and interpretability of Logistic Regression make it ideal for use cases where understanding the model’s decision-making process is crucial.

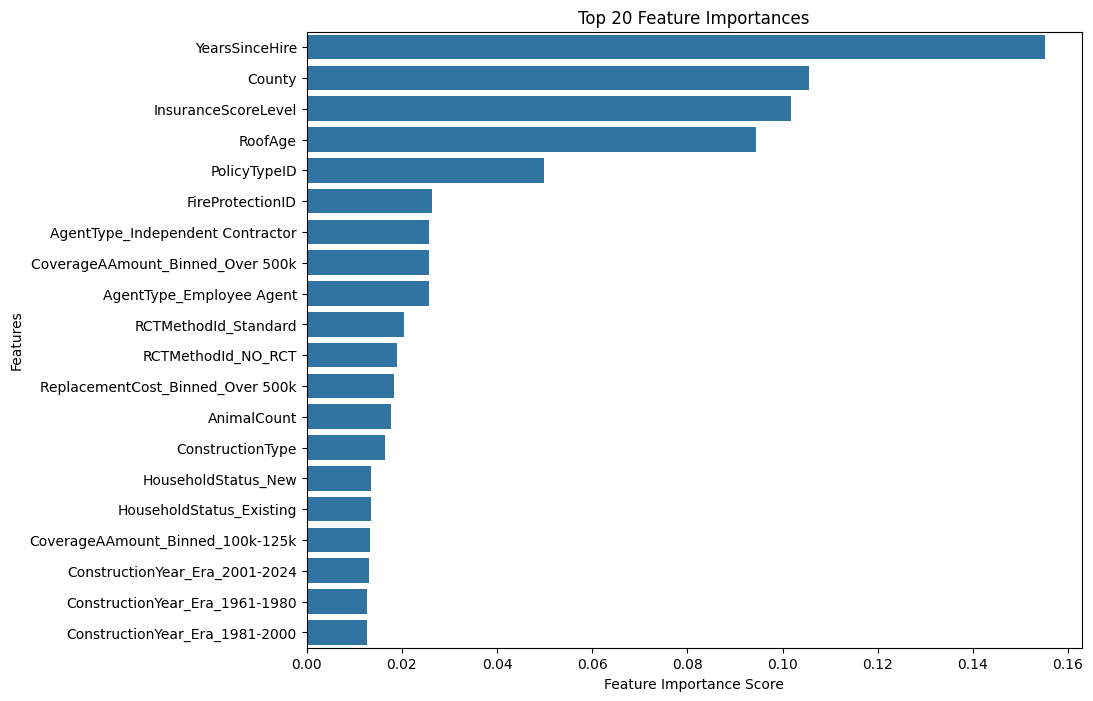


We also identified the most influential features for predicting the target variable 'Target\_Changes'. This model evaluation was aimed at understanding the relative importance of various features in the dataset, which could provide insights into factors that significantly affect the prediction outcomes.

The bar chart illustrates the top 20 features in terms of their importance, as derived from the RandomForest model. Notably, 'YearsSinceHire' emerges as the most significant predictor, suggesting that the duration an agent has been employed is highly indicative of the outcome. This is followed closely by 'County' and 'InsuranceScoreLevel', which indicate that geographical and insurance-related metrics are also crucial determinants.

Other important features include 'RoofAge', 'PolicyTypeID', and 'FireProtectionID', which are intuitively relevant in an insurance context, reflecting property conditions and policy specifics that could impact risk assessments. The presence of features such as 'AnimalCount' and different eras of 'ConstructionYear' (e.g., 2001-2024, 1961-1980, 1981-2000) further underlines the role of property characteristics and the time-related aspects of the insured properties.

The visualization and the corresponding analysis of these top features allow stakeholders to better understand the drivers behind the model’s predictions and to focus on key areas for policy adjustment, risk assessment enhancements, or further detailed analytical studies. This feature importance analysis is not only crucial for refining the predictive model but also offers actionable insights for strategic decision making in insurance policy management and risk evaluation.



## **CONCLUSIONS**

As the dilemma we stated in the first page, limited workforce and the property uncertain risk level are crucial for insurance company. Under this situation, optimizing the house inspection process to adapt the evolving economic conditions and inflationary pressures can play the role in saving the cost and improving the efficiency of inspection.

The current algorithm from the company applied 2 elements to determine whether the house need to be inspected, the first one is if the agent is on the waive list, the second is if the Coverage A value is in the range of 125,000 to 500,000. This method can target some house exist problems during inspections but the accuracy is not meet expectation as it may lack for some important elements related to the risk. Our new predictive model detected ten most influential factors to the risk level. The first three are about agent information, the years of experience of an agent is the most important factor, which means judgements made by experienced agents are crucial for risk mitigation. The house roof age, coverage A amount larger than 500k, fire protection ID, animal count, and construction type are important factors which are collected from the inspection table, which means changes or issues happened on these labels imply a significant rise in risk and requires attention. As these changes can be tracked on the inspection table, the insurance company could focus more on these variables when check the forms and house condition so that can reduce unnecessary inspection, saving time and costs and increase efficiency.

In order to quantify our optimization results, we generate a cost analysis for the two models. Defined inspection that happened but found nothing changed as redundant, the current algorithm has 1324 redundant cases yearly in the dataset, whereas our new model only has 270 annually. The company can reduce 38% of redundant cases and save $421k per year with the average cost per inspection is around $400 in America. In conclusion, this new predictive model can beneficial the business through increase detect accuracy, reduce the risk for prediction errors and financial pressures and gain higher customer satisfaction due to the less disturbing.

However, there are still some drawbacks exist in our model which may influence the accuracy. The first point is that it is hard for now to implement inflation rate in the model. Although the mean of the coverage A value was increased a lot in our tracked period due to the inflation, this increase rate is not a constant number and changed with complex market factors, it is hard to calculate or predict a precise number. If we have an accurate model to predict this rate in the future, this work can be considered in our model and the accuracy will be ideal. The second thing is the limitation of the data scope. Our model is based on the exist data, however, the data may be out of data and the market trend may be very different from now. If some dynamic factors which can track the market situation can be added in our model, the model can be automatically updated and its application can be much stronger.

## **REFERENCES**

Ahn, J. J., Byun, H. W., Oh, K. J., & Kim, T. Y. (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. Expert Systems with Applications, 39(9), 8369-8379.

Ogunnaike, R. M., & Si, D. (2017). Prediction of insurance claim severity loss using regression models. In Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, MLDM 2017, New York, NY, USA, July 15-20, 2017, Proceedings 13 (pp. 233-247). Springer International Publishing.

Chauhan, Abhishek and Arora, Mridul and Jain, Rachna and Nagrath, Preeti, Evaluation of Real Estate Appraisal Using Ensemble Methods (December 12, 2020). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021.

Schmitt, J., Bönig, J., Borggräfe, T., Beitinger, G., & Deuse, J. (2020). Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing. Advanced engineering informatics, 45, 101101.

Taha, A., Cosgrave, B., & Mckeever, S. (2022). Using feature selection with machine learning for generation of insurance insights. Applied Sciences, 12(6), 3209.

Seger, C. (2018). An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing.

Kumar, S. P. (2013). Prediction of consumer purchase decision using demographic variables: A study with reference to premium car. IOSR Journal of Business and Management, 12(5), 117-120.

Santhanam, Ramraj & Uzir, Nishant & Raman, Sunil & Banerjee, Shatadeep. (2017). Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets.

## **AI RESEARCH TOOL REFLECTION**

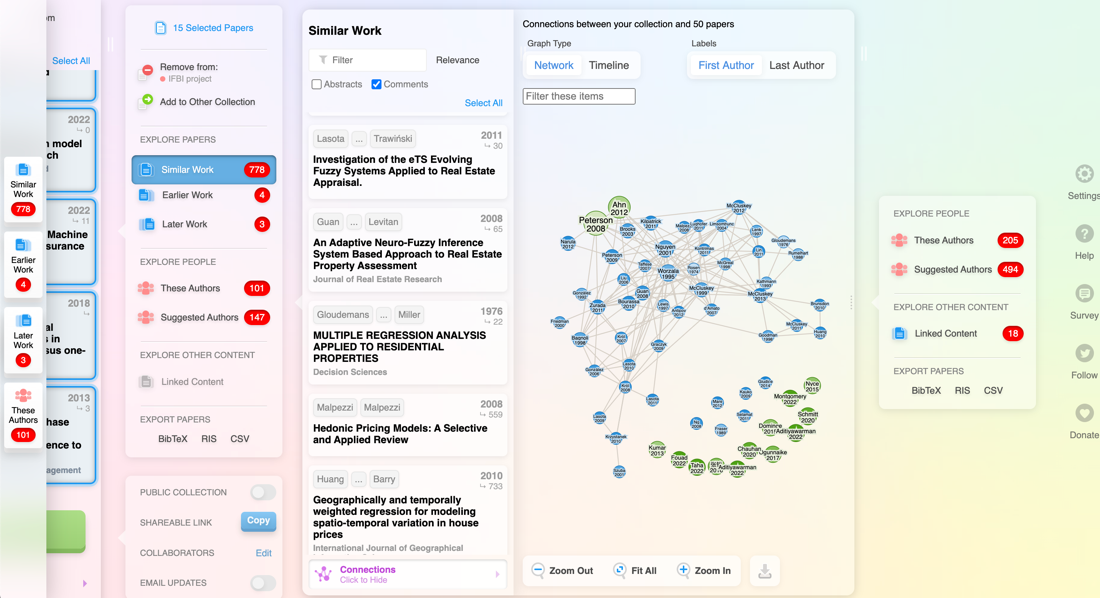
Literature review serves as a cornerstone in the realm of academic inquiry, providing a comprehensive synthesis of existing knowledge and insights relevant to a particular research area. However, the sheer volume and diversity of available literature often pose significant challenges for researchers seeking to extract meaningful insights. In this context, the integration of Artificial Intelligence (AI) tools emerges as a transformative solution, revolutionizing the traditional paradigms of literature review.

In our research methodology, we leverage AI to enhance the efficiency of our literature review process.

Initially, we use AI as a helper to guide us in selecting appropriate and comprehensive keywords for article searches.

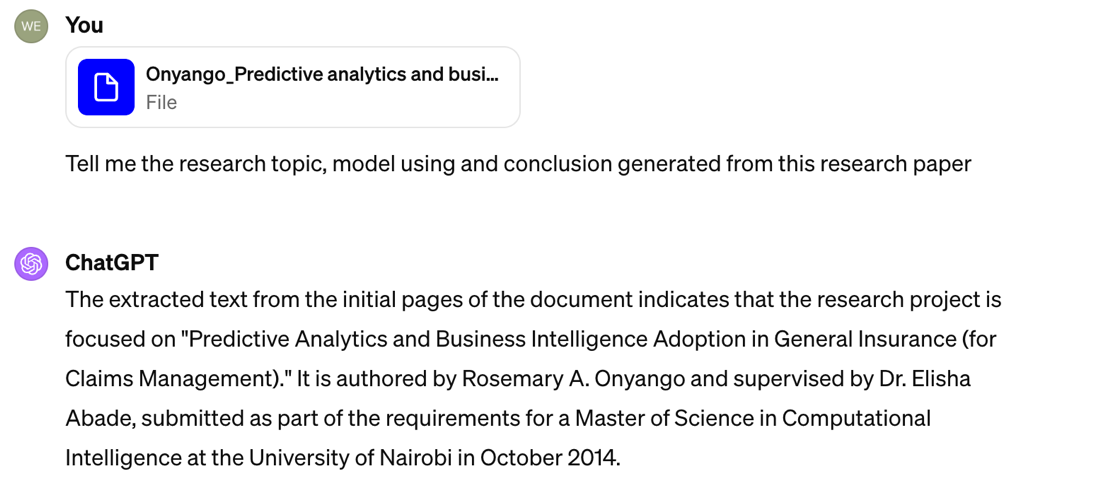
## 

In addition to the initial keyword searches, AI tools enable us to generate additional ideas and broaden the scope of our search, facilitating the discovery of high-quality articles for reference. Tools such as **Research Rabbit** was used extensively and helped us find high quality papers relevant to our problem statement with ease.

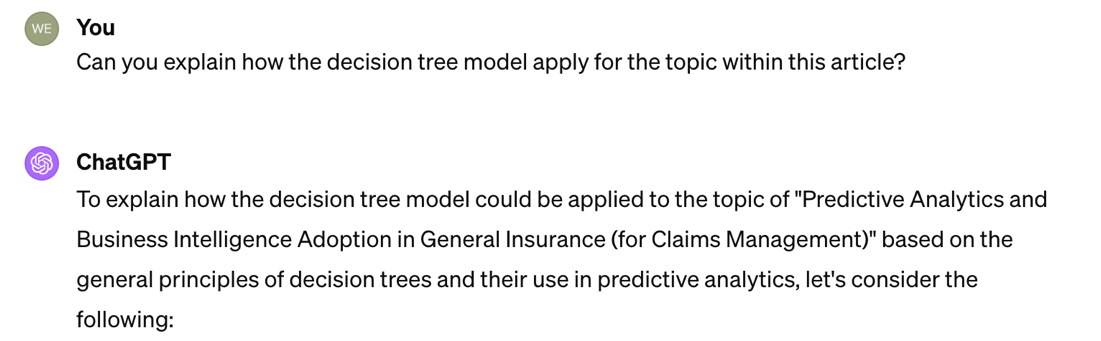


Secondly, when confronted with a vast number of articles covering similar topics or featuring valuable models, the challenge of reviewing each one thoroughly within time constraints becomes evident. To address this issue, we enlist AI assistance to summarize key information such as the research topic, methodology, and utilized models.

By sharing the files or links of relevant articles, AI aids in generating concise summaries, allowing us to discern the direction of our research and saving considerable time on search and review processes.



When encountering new concepts, AI serves as a valuable tool in expediting the learning process and facilitating the integration of these concepts into our research topic. It acts as a proficient assistant in comprehending articles authored by knowledgeable predecessors, thereby reducing the time required to familiarize oneself with a new domain.



## 

Lastly, AI tools play a crucial role in refining the language and structure of our literature review summaries. They ensure the correctness of our words and enhance the professionalism of our sentences, thereby benefiting both writers and readers alike.

Nevertheless, one must exercise caution and mindfulness when relying on AI for assistance. While these tools offer efficiency and accuracy, their limitations should be acknowledged. AI may occasionally struggle to fully comprehend questions or provide answers that deviate from the intended direction. Therefore, it is imperative to critically evaluate the output generated by AI tools and avoid overdependence on them.

## **APPENDEIX**

You can put supporting things here. Sometimes people will put their data dictionary table here and refer to it in the Data section.

Data Dictionary:

Table 1: Data used in study

|  |  |  |
| --- | --- | --- |
| Inspection | | |
| Variable | Type | Description |