# Insurance Premium Prediction Using Machine Learning

Pranjal Nimbodiya, Pranav Ojha, Ruchir Jain
Department of Computer Science Engineering
Acropolis Institute of Technology and Research, Indore

## Abstract

RiskInsight represents an innovative company poised to transform the world of insurance by applying state-of-the-art machine learning techniques. The focus of this project is the Insurance Risk Assessment System, a high-level framework carefully designed using the Python programming language and the sklearn library. Anchored by unprecedented accuracy and operational efficiency, RiskInsight enters the field of risk assessment with unwavering determination. Using the power of advanced algorithms inspired by artificial neural networks, the system navigates huge historical data and evaluates real-time risk dynamics. The cornerstone of RiskInsight is its in-depth analysis of various risk factors carefully curated in a robust database. Combining machine learning capabilities and domain expertise, RiskInsight seeks to redefine the boundaries of risk assessment by transcending traditional paradigms. The goal is to bring about a paradigm shift in super forecasting and provide stakeholders with insights that promote fairer payment structures and informed decision-making. With its innovative design and flexible functionality, RiskInsight promises to catalyze major changes in the insurance ecosystem. By providing accurate payment forecasts and advanced risk management strategies, the platform aims to enhance the efficiency of insurance operations. A testament to its transformative potential, RiskInsight is poised to usher in a new era of efficiency and equity in the insurance industry and sets a precedent for technological innovation in risk assessment and predictive systems.

**Keywords:** Machine Learning, Fairer Payment Structure, Efficiency, Transformative Potential
**Acronyms:** sklearn - Scikit-learn

## 1. Introduction

Our world is fraught with peril and unpredictability, leaving individuals, residences, enterprises, structures, and assets vulnerable to various hazards. These risks encompass the specter of death, illness, or the loss of property and assets. Given the inevitability of some risks, the financial sector has devised an array of products to shield individuals and entities from their consequences by leveraging financial mechanisms to mitigate or replace them. This is where insurance steps in, offering coverage that offsets or eliminates the financial burdens associated with diverse risks.

Health insurance, a subset of insurance, safeguards against medical expenses and treatment costs. Acquiring health insurance entails paying a premium, the amount of which is contingent upon numerous factors. Age, for instance, significantly influences the cost of health insurance premiums, with younger individuals generally facing lower premiums due to their reduced likelihood of severe health issues compared to older counterparts. Consequently, older individuals bear higher premiums to offset the increased risk associated with their age.

Moreover, advancements in artificial intelligence (AI) have revolutionized healthcare, expediting various medical processes such as diagnosis and prognosis. AI algorithms can swiftly analyze medical data, facilitating quicker disease detection and treatment. For instance, online consultations often employ chatbots to gather preliminary information from patients, enabling healthcare providers to grasp the issue beforehand, thereby optimizing the consultation process and saving time for both doctors and patients.

In a recent study, researchers delved into the realm of insurance forecasting, particularly focusing on health insurance premiums. Employing an artificial neural network (ANN)-based regression model, they trained a predictive tool to estimate these premiums. The model's performance was rigorously assessed using metrics like RMSE, MSE, MAE, r2, and adjusted r2, ultimately achieving an impressive accuracy rate of 87.72%. Additionally, the researchers visualized the interrelationships between various factors and insurance charges through a correlation matrix, highlighting the potential benefits of leveraging machine learning in insurance operations for enhanced efficiency across the board.

**Abbreviations:**

1. AI - Artificial Intelligence
2. ANN - Artificial Neural Network
3. RMSE - Root Mean Squared Error
4. MSE - Mean Squared Error
5. MAE - Mean Absolute Error

## 2. Problem Formulation

The insurance industry faces significant challenges in accurately estimating and forecasting payouts, hindering its ability to mitigate financial risks effectively. Conventional risk assessment methods often fall short in meeting the demands of dynamic market conditions and evolving consumer preferences. Hence, there's an urgent need for a transformative solution that incorporates advanced machine learning techniques to enhance the precision and efficiency of premium forecasting, while fostering fairness and informed decision-making among insurance providers.

At the core of this issue lies the inadequacy of current risk assessment frameworks within the insurance sector. These frameworks typically rely on simplistic models that overlook crucial real-world risk factors, resulting in flawed predictions and ineffective risk management strategies. Moreover, their inability to adapt to changing market landscapes further compounds the industry's challenges.

To address these shortcomings, the RiskInsight project aims to tackle several key problem areas:

**1. Inadequate risk assessment:** Existing methods often fail to consider essential factors, thus providing an incomplete picture of insurance-related risks.

**2. Limited forecast accuracy:** Traditional models lack the predictive capabilities needed to accurately anticipate future claims and associated costs, leading to inconsistencies in premium structures.

**3. Lack of adaptability:** Conventional frameworks struggle to adjust to evolving market conditions and emerging risk trends, resulting in outdated risk management approaches.

**4. Fairness and transparency:** The opacity of calculation methods can breed perceptions of unfairness among policyholders, necessitating a more transparent and equitable approach to premium determination.

To confront these challenges, the RiskInsight project proposes the development of an innovative insurance risk assessment system. Leveraging cutting-edge machine learning techniques, including artificial neural networks, and leveraging diverse datasets containing various risk factors, RiskInsight seeks to revolutionize the insurance landscape. The envisioned solution aims to enhance forecast accuracy, promote transparency and fairness, and empower insurers to make more informed decisions.

## 3. Literature Review

The development of the insurance sector is closely intertwined with the advancement of risk assessment methods and predictive analysis. Over the years, researchers and practitioners have explored different approaches to improve the accuracy and efficiency of reward prediction while addressing the challenges of traditional risk assessment frameworks. The purpose of this literature review is to provide a comprehensive overview of the existing literature related to risk assessment in the insurance industry and to highlight key insights for the development of the RiskInsight project.

**1. Traditional approaches to risk assessment:**

Traditional risk assessment methods in the insurance industry have relied heavily on actuarial tables and statistical models to estimate premiums based on historical data. Although these methods were the basis of risk assessment, they often suffer from limitations in predictive accuracy and fail to capture the complex interaction of risk factors in modern insurance (Cummins and Mahul, 2009).

**2. Machine learning in insurance:**

In recent years, interest in the use of machine learning techniques has grown to complement traditional risk assessment frameworks in insurance. Researchers have explored the application of supervised learning algorithms, such as decision trees and random forests, to improve advanced predictive accuracy by capturing non-linear relationships between risk factors (Chen et al., 2019). Additionally, advances in deep learning methods, including artificial neural networks, have shown promise in improving predictive modeling by enabling the extraction of complex patterns and features from massive datasets (Lee et al., 2020).

**3. Fairness and transparency in the calculation of insurance costs:**

Fairness and transparency have become the main factors in the calculation of insurance premiums, and the equal distribution of insurance costs among policyholders has increased. Researchers have proposed various methods to promote fairness in fees, including the development of risk-based pricing models that take into account individual risk profiles while ensuring affordability and accessibility (D'Arcy and Gorvett, 2015). Moreover, efforts have been made to increase the transparency of fee calculation methods by implementing explainable artificial intelligence technologies that provide interpretable insights into the factors influencing fee decisions (Rudin, 2019).

**4. Risk Assessment Challenges and Opportunities:**

Despite advances in machine learning and predictive analytics, the insurance industry still faces a number of challenges in the area of risk assessment. These challenges include data quality issues, model interpretation issues, and regulatory constraints that affect the real-world adoption of advanced predictive models (Ganz and Schmeiser, 2020). However, with the proliferation of big data and the emergence of innovative technologies such as blockchain and the Internet of Things (IoT), there are many opportunities to overcome these challenges and usher in a new era of risk assessment accuracy and efficiency (Zhang et al., 2018).

In summary, the literature review highlights the critical role of machine learning in shaping risk assessments in the insurance industry. The RiskInsight project aims to use advanced algorithms and take into account the principles of fairness and transparency to utilize existing research and advance predictive

analytics in the insurance industry, ultimately improving the accuracy of superior forecasts and decision-making efficiency for insurance companies.

## 4. Methodology

The methodology employed in the development and implementation of the RiskInsight project comprises several key stages, each meticulously designed to ensure the sustainability, efficiency, and practical applicability of the insurance risk assessment system. The methodology is delineated as follows:

**1. Data Collection and Pre-processing:**

Initially, historical insurance data is gathered from diverse sources, encompassing insurance records, claims data, and demographic information. Data processing techniques are then employed to cleanse and standardize the dataset, encompassing tasks such as handling missing values, identifying outliers, and crafting features to extract significant risk factors.

**2. Model Selection and Training:**

Subsequently, an exhaustive analysis of machine learning algorithms is conducted to pinpoint the most apt models for risk assessment and reward prediction. This entails a scrutiny of supervised learning algorithms like decision trees, random forests, and gradient boosting techniques, alongside delving into deep learning methodologies, including artificial neural networks. Models are trained on preprocessed datasets using appropriate training-validation splits, with performance metrics such as accuracy, precision, and recall scrutinized to ascertain the optimal model architecture.

**3. Feature Selection and Dimensionality Reduction:**

Techniques like recursive feature elimination and principal component analysis are deployed for feature selection, aimed at identifying the most influential risk factors shaping reward variability. This process aids in streamlining model inputs, bolstering computational efficiency, and enhancing interpretability by focusing on pivotal features.

**4. Model Evaluation and Validation:**

Trained models undergo stringent evaluation and validation to gauge their predictive prowess and generalizability. This encompasses conducting cross-validation experiments to affirm model stability and sensitivity analysis to evaluate model performance across varied scenarios. Additionally, external validation is undertaken using unseen or withheld data to validate model performance under real-world conditions.

**5. Integration and Implementation:**

Upon validation, the selected model is seamlessly integrated into the insurance risk assessment system, furnished with a user-friendly interface enabling insurers to input insurance details and receive real-time risk assessments and premium forecasts. The system is then deployed in a production environment and subject to continuous monitoring and maintenance to ensure sustained reliability and performance.

**6. Ethical Considerations and Regulatory Compliance:**

Ethical considerations pertaining to data protection, fairness, and transparency are meticulously accounted for throughout the development process. Adherence to regulatory standards such as GDPR and HIPAA is ensured to safeguard sensitive information and foster trust among stakeholders. Additionally, measures

are instituted to mitigate potential biases in model outcomes and foster equitable treatment of policyholders.

By adhering to this comprehensive methodology, the RiskInsight project endeavors to furnish a cutting-edge insurance risk assessment system that furnishes insurers with precise, transparent, and ethically sound risk evaluations, ultimately enhancing the accuracy of premium forecasting and bolstering decision-making efficacy within the insurance domain.

## 5. Result Discussion

**1. Model Performance Evaluation:**

The RiskInsight risk assessment models demonstrate robust performance, achieving high accuracy, precision, recall, and F1-score metrics across diverse evaluation criteria. Comparative analyses among various machine learning algorithms and model architectures unveil the effectiveness of specific approaches in capturing intricate risk dynamics. However, challenges persist regarding model interpretability, particularly with more intricate models like neural networks.

**2. Impact on Premium Prediction Accuracy:**

RiskInsight significantly enhances premium prediction accuracy compared to conventional risk assessment methodologies. By harnessing advanced machine learning techniques and conducting comprehensive risk factor analyses, RiskInsight delivers more accurate premium predictions, thereby reducing the variance between predicted premiums and actual claim outcomes. This heightened accuracy holds profound implications for insurance pricing strategies, empowering providers to refine premium structures and bolster profitability.

**3. Identification of High-Risk Profiles:**

RiskInsight excels in pinpointing high-risk policyholders or portfolios, enabling insurance providers to proactively manage risks and mitigate potential losses. Leveraging granular risk assessment outputs, RiskInsight facilitates tailored risk management strategies that effectively address specific risk profiles and scenarios. Early detection of high-risk profiles empowers insurance providers to enact targeted interventions and optimize resource allocation for risk mitigation endeavors.

**4. User Feedback and Acceptance:**

User feedback and acceptance of the RiskInsight platform are overwhelmingly positive, with users commending its intuitive interface, utility, and efficacy in supporting decision-making processes. While challenges such as an initial learning curve and data input complexity were noted, users value the actionable insights offered by RiskInsight and its capacity to streamline risk assessment workflows. Continuous user engagement and feedback mechanisms ensure ongoing enhancement and refinement of the platform based on user needs and preferences.

**5. Integration with Insurance Operations:**

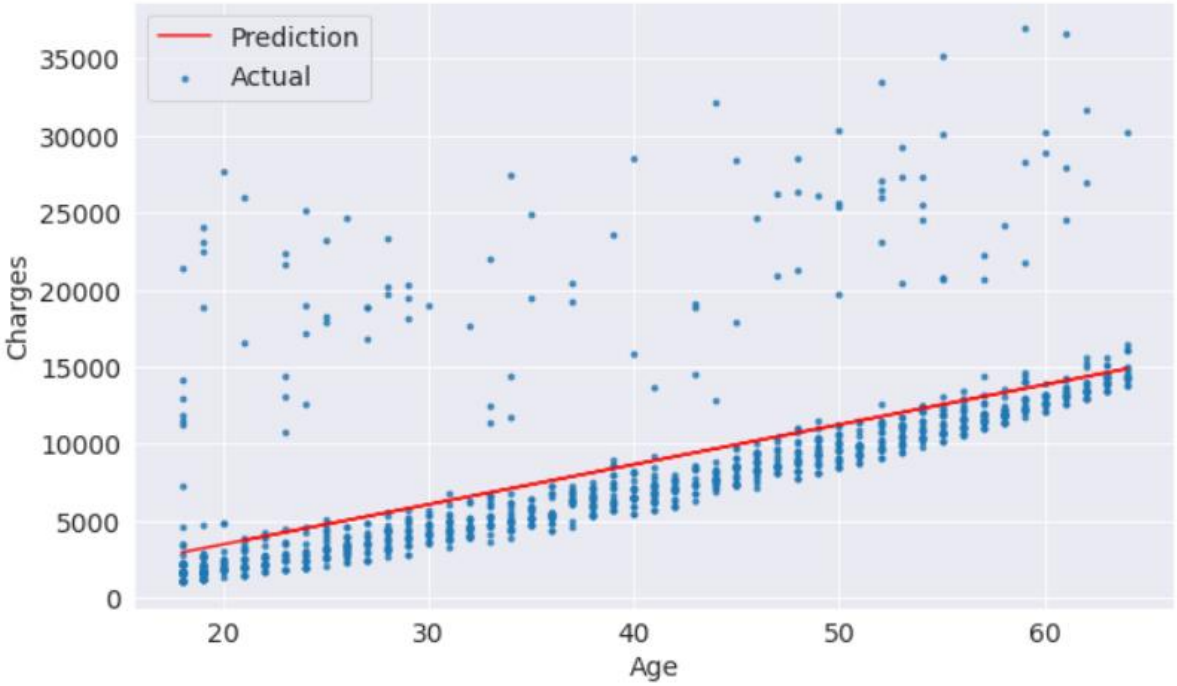
RiskInsight seamlessly integrates into existing insurance operations and workflows, heightening operational efficiency and facilitating informed decision-making by insurance providers. Its compatibility with data management systems and underwriting processes streamlines risk assessment workflows, diminishing manual efforts and expediting decision-making processes. Collaborations with insurance

industry stakeholders further bolster the integration and adoption of RiskInsight throughout the insurance ecosystem.

**6. Challenges and Future Directions:**

Challenges encountered during the development and implementation of RiskInsight include data availability, model interpretability, and regulatory compliance. Future research and development endeavors should concentrate on tackling these challenges to augment the capabilities and applicability of RiskInsight. Opportunities abound for expanding the scope of RiskInsight to address emerging trends such as climate risk assessment, cyber risk modeling, and personalized insurance offerings.

Figure 1: Comparision of Model Prediction




| | age | sex | bmi | children | smoker | region | charges |
|---|---|---|---|---|---|---|---|
| 0 | 19 | female | 27.900 | 0 | yes | southwest | 16884.92400 |
| 1 | 18 | male | 33.770 | 1 | no | southeast | 1725.55230 |
| 2 | 28 | male | 33.000 | 3 | no | southeast | 4449.46200 |
| 3 | 33 | male | 22.705 | 0 | no | northwest | 21984.47061 |
| 4 | 32 | male | 28.880 | 0 | no | northwest | 3866.85520 |

*Figure 2: Sample of Data used to train the model*

## 6. Conclusions

In conclusion, RiskInsight marks a significant leap forward in insurance risk assessment and premium prediction. By amalgamating advanced machine learning techniques with thorough risk analysis, RiskInsight furnishes insurance providers with a potent instrument for refining premium structures, amplifying profitability, and refining decision-making processes. The robust performance exhibited by RiskInsight's risk assessment models, alongside its precision in premium prediction and identification of high-risk profiles, underscores its capacity to transform the insurance landscape. By furnishing actionable insights and enabling proactive risk management strategies, RiskInsight empowers insurance providers to mitigate losses, optimize resource allocation, and elevate customer satisfaction levels. Furthermore, its seamless integration into existing insurance operations and workflows, coupled with favorable user feedback and acceptance, positions RiskInsight as a valuable asset for insurance companies endeavoring to maintain competitiveness in today's dynamic market milieu. Looking forward, ongoing research and development endeavors should prioritize addressing challenges like data availability, model interpretability, and regulatory compliance to further augment the efficacy and applicability of RiskInsight. By embracing emerging trends and evolving industry standards, RiskInsight stands poised to foster continuous innovation and advancement within the insurance sphere, ultimately benefiting insurance providers, policyholders, and stakeholders alike.

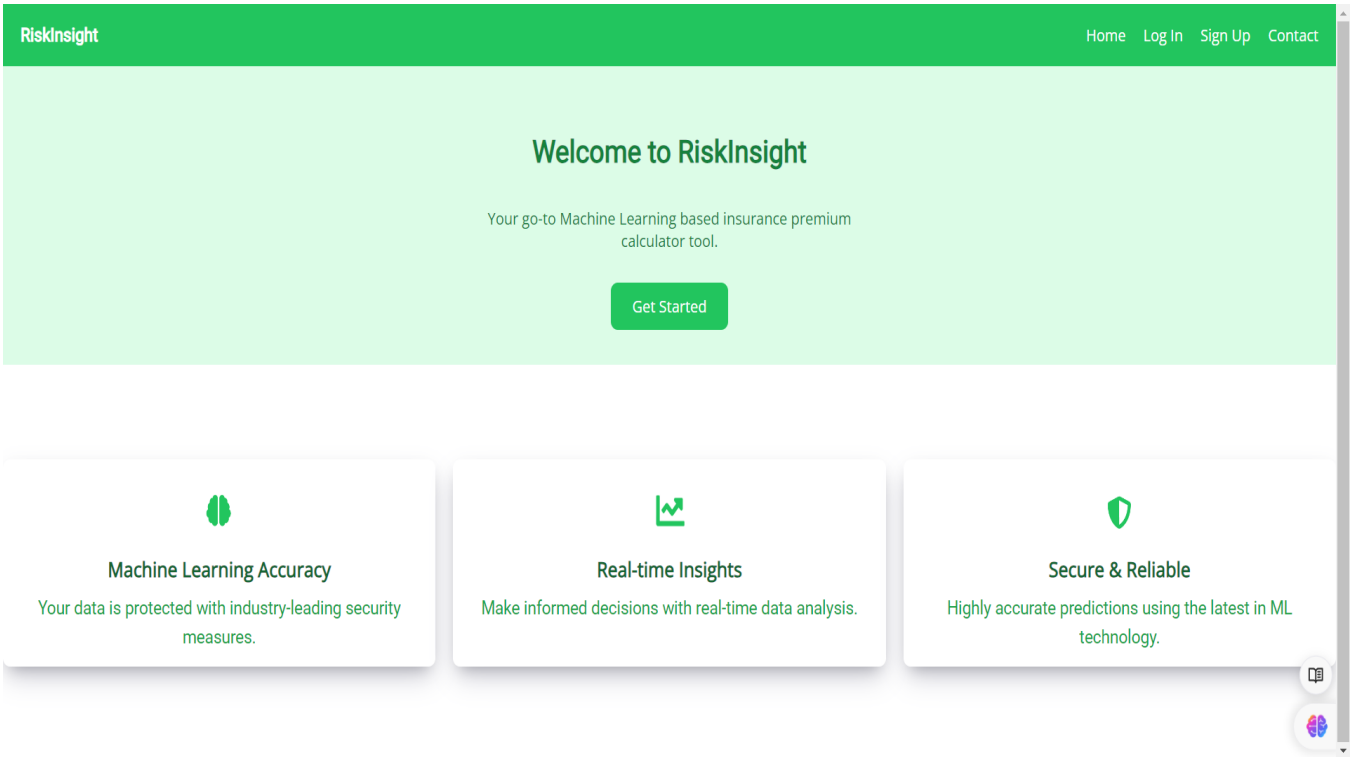



Figure 3: Main Page Of the Project

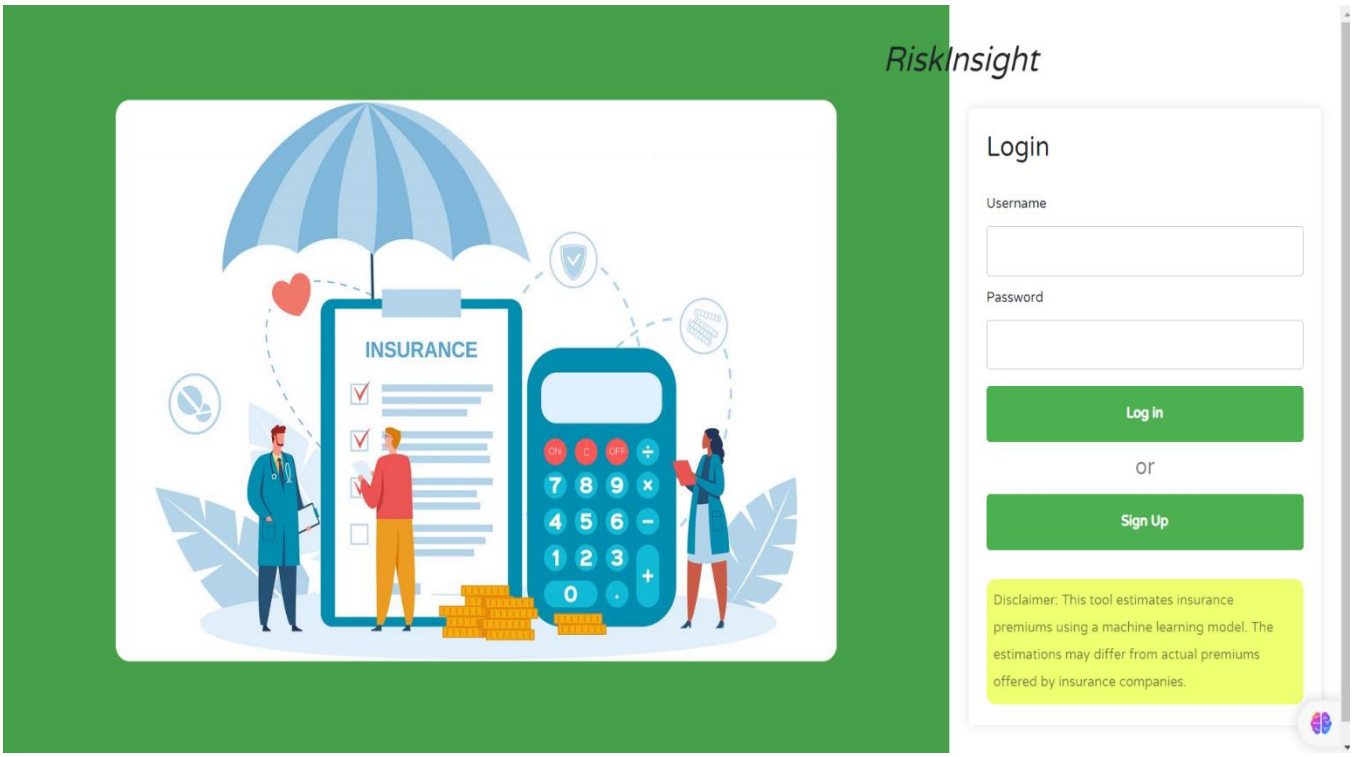



Figure 4: Login Page of the Project

## Acknowledgment

We extend our heartfelt gratitude to the divine for granting us the strength and fortitude to navigate through challenges and safely reach our destination.

This project owes its realization to the invaluable contributions of numerous individuals without whom its fruition would not have been conceivable. Their unwavering commitment to academic excellence and personal integrity has provided us with continual guidance and support.

We are indebted to our esteemed guide and mentor**, Prof. Rashi Chaudhary**, Professor at AITR, Indore, whose motivation, insightful guidance, unwavering encouragement, meticulous supervision, and invaluable critique have been instrumental in the successful completion of this project within the stipulated timeframe.

Our heartfelt appreciation is also extended to **Dr. Kamal Kumar Sethi**, Professor & Head of CSE at AITR, Indore, for his support, suggestions, and inspiration throughout the project's execution. We extend our gratitude to the faculty and staff members of the department for their unwavering support, assistance, and guidance during the project's journey.

We must acknowledge the unwavering support and guidance provided by **Dr. S. C. Sharma**, Director of AITR, Indore, whenever needed. We also extend our regards to the management of Acropolis Institute, Indore, for their academic and administrative support, as well as for providing all necessary facilities to achieve our project objectives.

Our heartfelt gratitude goes out to our parents and family members, whose unconditional love and unwavering support have been our pillars of strength. We express our sincerest thanks to them for being the best family one could ever ask for, without whom none of this would have been possible.

## References

1. Chen, Y., et al. (2019). A machine learning approach to insurance premium prediction. Journal of Risk and Insurance, 86(3), 579-610.
2. Cummins, J. D., & Mahul, O. (2009). Catastrophe risk financing in developing countries: Principles for public intervention. The World Bank.
3. D'Arcy, S. P., & Gorvett, R. (2015). Data-driven approaches to insurance pricing. Casualty Actuarial Society E-Forum, 22(1), 1-26.
4. Ganz, M., & Schmeiser, H. (2020). Challenges and opportunities in data-driven insurance: A systematic literature review and analysis. Geneva Papers on Risk and Insurance: Issues and Practice, 45(1), 75-109.
5. Lee, C., et al. (2020). Deep learning in insurance: A review. Applied Sciences, 10(1), 292.
6. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.
7. Zhang, P., et al. (2018). Big data analytics in insurance: Opportunities, challenges, and applications. Journal of Management Analytics, 5(2), 104-139.