# Predicting Risk for New Insurance Using Customer Information

**A Major Project Synopsis Submitted to**



**Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal Towards Partial Fulfillment for the Award of**

**Bachelor of Engineering (Computer Science Engineering)**

## Under the Supervision of Prof. Narendra Pal Singh Associate Professor (P)

**Submitted By**

1. **Durgesh Sharma (0827CS181072)**
2. **Harshala Gaikwad (0827CS181087)**
3. **Kartavya Verma (0827CS181103)**

**Department of Computer Science Engineering Acropolis Institute of Technology & Research, Indore Jan-June 2022**

Project Synopsis/Proposal Guidelines

### Abstract

* + Insurance policies involving factors with greater risk of claims are charged at a higher rate. With much information at hand, insurers can evaluate risk of insurance policies at much higher accuracy.
  + With the increase in the amount of data and advances in data analytic, the underwriting process can be automated for faster processing of applications. This project aims at providing solutions to enhance risk assessment among life insurance firms using predictive analytic.

### Introduction of the Project:

* + Risk assessment is a crucial element in the life insurance business to classify the applicants. Companies perform underwriting process to make decisions on applications and to price policies accordingly.
  + Risk profiles of individual applicants are thoroughly analyzed by underwriters, especially in the life insurance business. The job of the underwriter is to make sure that the risks are evaluated, and premiums as accurately as possible to sustain the smooth running of the business.
  + The use of advanced data mining techniques to improve decision making has already taken root in property and casualty insurance as well as in many other industries.
  + Individual life insurance organizations still rely on the conventional actuarial formulas to predict mortality rates and premiums of life policies. We have concentrated on data mining techniques to detect risk among insurance firms, which is a crucial issue due to the companies facing great losses.
  + The predicted risk level of the insurance holder helps insurance to calculate the premium and display it to the policy holders

### Objective:

* + The purpose of this project is to apply predictive modeling to classify the risk level based on the available past data in the life insurance industry and customer information and recommend the most appropriate model to assess risk and provide solutions to refine underwriting processes. These risk further help the insurer to evaluate the premium for insurance.
  + Companies that underwrite life insurance policies have to evaluate applications carefully. The payouts from life insurance claims are very high relative to the insurance premiums that companies collect from an individual customer. So, this project will surely reduce the high cost pay-outs life insurance claims.

### Scope:

* + Predictive modeling utilizing learning algorithms can give the eminent contrast in the manner which business is done as compared with the traditional strategies.
  + With information expository arrangements, the work should be possible quicker and with better outcomes.
  + Future work identifies with the more top to bottom investigation of the issue and new techniques to manage specific systems.
    - Client division is the division of the informational index into gatherings with comparable credits can be executed to section the candidates into gatherings with comparable qualities dependent on the characteristics present in the data-set.
    - For instance, comparable work history, protection history and restorative history.
    - Following the gathering of the candidates, prescient models can be executed to add to an alternate information digging approach for the disaster protection client informational collection.

### Study of Existing System:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Existing system/website/software** | **Features** | **Disadvantages** | **Limitations/Gaps** |
| 1. | Predictive Modeling for Life Insurance | In these different ways of data mining and multivariate analytic techniques was described that can be used to improve decision making processes in such functions as life insurance underwriting and marketing | Risk factor and mean standard evaluation is not done. | Data used was limited. |
| 2. | Insurance Claim Analysis Using Machine Learning Algorithms | In this system, several machine learning techniques to analysis the insurance claims efficiently and compare their performances using various metrics. | Security risk with a lot of confidential data of the user. | Data security. |
| 3. | Predictive Modeling for Life Insurance | This system illustrates the general processes that can be used to implement predictive modelling in life insurance underwriting and marketing. | User data collection was not up-to the mark. | Unpredictability  of the result. |

1. Project Description:
   * In this project, an insurance risk predictive modeling approach is discuss for predicting risk for new insurance using customer information.
   * This approach is about RME-EP (Rule-based Model Evaluation with Event Processing) which is an amalgam of predictive modeling and (forward chaining) rule engine.
   * It uses Deep Learning approach for improving the accuracy of system. RME-EP can be used to combine several predictive models into a single model, producing combined predictions such as maximum, minimum, average, etc.
   * It uses XGBoost classifier which results in better accuracy, high precision, increase in speed and good performance of the overall project.
   * In addition, it can be used to classify combined predictions into classes such as "Very high risk", "High risk", "Medium risk", "Low risk", etc.
   * Instead of treating all customers equally, managers have come to understand that it is more effective to develop customer-specific strategies.
   * By using customer information contained in databases, companies can invest in the customers that are potentially valuable for the company, but also minimize their investments in non-valuable customers.

### Methodology/Planning of the Project Work:

* + This approach involves the gathering of knowledge from online databases. The project paradigm deals with a positivist approach because it is especially a predictive study involving the utilization of machine learning algorithms to support the research objectives.
  + **Step 1:** Data collection and data-set preparation This will involve collection of customer information containing data-set that is real world data-set so this raw data-set needs pre-processing and feature extraction by noise reduction, feature extraction and data normalization so that this preprocessed data can be used on the proposed model.
  + **Step 2:** Developing An RME-EP (Rule-Based Model Evaluation with Event Processing) model for Predicting Risk for New Insurance Using Customer Information In this step a Deep Learning based approach i.e., RME-EP (Rule-based Model Evaluation with Event Processing) which is an amalgam of predictive modeling and (forward chaining) rule engine is developed for predicting risk for new insurance using customer information.
  + **Step 3:** Training and experimentation on data-sets The RME-EP (Rule-based Model Evaluation with Event Processing) predictive model that is Deep Learning based approach will be trained on the data-set to do the predicting risk for new insurance using customer information and notify the insurance company about the risk factors of customer by dividing it into classes such as "Very high risk", "High risk", "Medium risk", "Low risk”.
  + **Step 4:** Deployment and analysis on real life scenario.The trained and tested Predicting Risk for New Insurance Using Customer Information model will be deployed in a real-life scenario made by the human experts & will be leveraged for further improvement in the methodology and will follow the above architecture.

### Expected Outcome:

* + The insurance companies are tremendously interested in the prediction of the future. Accurate prediction gives a probability to decrease financial loss for the company. The insurers use rather complex methodologies for this purpose. The algorithms involve detection of relations between claims, implementation of high dimensionality to reach all the levels, detection of the missing observations, etc. In this way, the individual customer’s portfolio is made. Forecasting the upcoming claims helps to charge competitive premiums that are not too high and not too low. It also contributes to the improvement of the pricing models. This helps the insurance company to be one step ahead of its competitor.
  + This approach wants to make insurance quicker and less labour intensive for new and existing customers to get a quote while maintaining privacy boundaries. By developing a predictive model that accurately classifies risk using a more automated approach, we can greatly impact public perception of the industry.

### Resource and limitations:

* + Insurance was always based on data analysis: accident statistics, policyholder’s personal information, as well as third-party sources help to group people into different risk categories, prevent fraud losses, and optimize expenses. The rapid movement towards the Digital Society opens new sources of information that can be used to create a complex behavioural pattern for each particular customer and precisely determine his or her risk class.
  + There are two new data sources:
    - Online behaviour – this includes social media activity, online shopping behaviour, browsing activity, etc.
    - Sensor data – from devices on the Internet of Things such as drones, Smart homes, cars.
  + Such personal data can complement the traditional sources used in insurance, generating real-time insights about a person's lifestyle and habits that can be used for competitive advantage.
  + Some of the challenges insurers typically encounter when adopting machine learning are:
    - Training requirements
    - Right data source
    - Difficulty in predicting returns
    - Data security

### Conclusion:

Consequently, it would upgrade the business by enabling quicker administration to client, in this manner expanding fulfillment and unwaveringness. Future work identifies with the more top to bottom investigation of the issue and new techniques to manage specific systems. Client division is the division of the informational index into gatherings with comparable credits can be executed to section the candidates into gatherings with comparable qualities dependent on the characteristics present in the data-set. For instance, comparable work history, protection history and restorative history. Following the gathering of the candidates, prescient models can be executed to add to an alternate information digging approach for the disaster protection client informational collection.

### References:

* + [U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, ―Critical analysis of Big Data challenges and analytical methods,‖ J. Bus. Res., 2017.](http://www.iosrjournals.org/iosr-jce/papers/Vol18-issue6/Version-5/U180605113119.pdf)
  + [Y. Joly et al., ―Life insurance: Genomic stratification and risk classification,‖ Eur. J. Hum. Genet., 2014.](https://www.researchgate.net/publication/317610964_Fire_Extinguishing_Robot)
  + Z. Ge, Z. Song, S. X. Ding, and B. Huang, ―Data Mining and Analytics in the Process Industry: The Role of Machine Learning,‖ IEEE Access, 2017.
  + A. C. Wuppermann, ―Private Information in Life Insurance, Annuity, and Health Insurance Markets,‖ Scand. J. Econ., 2017.
  + N. Boodhun and M. Jayabalan, ―Risk prediction in life insurance industry using supervised learning algorithms,‖ Complex Intell. Syst., 2018.
  + Mishr K (2016) Fundamentals of life insurance theories and applications. In: 2nd ed, Delhi: PHI Learning Pvt Ltd.
  + Wuppermann A (2016) Private information in life insurance, annuity and health insurance markets. Scand J Econ 119:1–45
  + Kajiamuller, ―The Identification of Insurance Fraud – an Empirical Analysis Working papers on Risk Management and Insurance‖ no: 137, June 2013.