HEALTH SCORE CALCULATOR FOR INSAURANCE COMPANIES

Report submitted by:

Abhishek Kumar

11-07-2023

ABSTRACT

Like the banking industry, the credit score of an individual or a small business is calculated based on financial parameters and past records of the individual or business. This credit score enables the individual or business to obtain loans from banks, while also allowing banks and financing companies to identify potential customers. A similar type of product, based on a person's past medical history and current health status, can be developed for insurance companies that offer life insurance coverage.

The person's current health status is determined by pathological tests for major life-threatening diseases, the environment in which they live, the type of occupation they engage in, and whether they smoke cigarettes and drink alcohol, or have any other kind of addiction. For calculating the health score of a person, we can create a website or an app that employs a machine learning model that is trained on a very large dataset related to the above-mentioned parameters, identifies the most important predictive parameter out of many parameters from pathology reports or from the past medical history that determines the probability of the person becoming ill in the future, and then converts that probability into a score that can be used by the insurer. Using machine learning, we can create a system similar to the credit score that can be a very beneficial product for insurance companies, allowing us to earn money whenever these companies utilise our services.

1. INTRODUCTION

Credit scoring has become an essential tool in the banking industry, enabling lenders to assess the creditworthiness of individuals and businesses based on financial parameters and historical records. Inspired by this concept, we propose the development of a health scoring system for insurance companies, specifically targeting life insurance coverage. The health score will provide insights into an individual's current health status and the probability of major life-threatening diseases. By leveraging machine learning algorithms and a comprehensive dataset, we aim to create a website or mobile application that generates health scores based on factors such as medical history, lifestyle choices, and environmental conditions. This innovative system will empower insurers to accurately assess health risks, enhance underwriting practices, and offer personalized life insurance coverage.

Drawing parallels to credit scoring, our health scoring system will utilize key health-related factors to evaluate an individual's health risk. Pathological test results, lifestyle choices, occupation, and other relevant parameters will be analyzed to determine the likelihood of future health complications. Machine learning techniques will be employed to train a model on a diverse dataset, identifying the most influential parameters for health scoring. This model will convert the probabilities into a comprehensive health score that insurance companies can use to assess risk, customize coverage, and calculate fair premiums. By adopting this data-driven approach, insurers can make informed decisions, optimize risk management, and provide tailored life insurance offerings.

The proposed health scoring system has immense potential for insurance companies, enabling them to leverage data-driven insights to enhance underwriting practices. It also benefits individuals seeking life insurance, as it provides a streamlined and non-invasive alternative to traditional medical examinations. This project aims to revolutionize the insurance industry's underwriting process by developing a robust health scoring system that empowers insurers and offers individuals a more accessible and personalized approach to life insurance coverage.

2. PROBLEM STATEMENT

The objective of this project is to develop a machine learning-based system that calculates a health score for individuals, considering various factors such as their medical history, current health status, lifestyle choices, and environmental conditions. This health score will serve as a predictive indicator of an individual's probability of developing major life-threatening diseases in the future. The system aims to assist insurance companies in assessing the health risk associated with potential policyholders and offer appropriate life insurance coverage accordingly.

3. MARKET AND CUSTOMER NEED

3.1. Insurance Companies:

→ Market Need: Insurance companies need an accurate and efficient method to assess the health risks associated with potential policyholders. Traditional underwriting processes often rely on subjective assessments, leading to potential inaccuracies and inconsistencies.

A health scoring system can provide insurers with a data-driven and objective approach to evaluate the health status of individuals, enabling better risk assessment and appropriate pricing of life insurance policies.

→ Customer Need: Insurance companies require a reliable tool that can effectively predict an individual's probability of developing major life-threatening diseases. By utilizing a health scoring system, insurers can better understand and manage their policyholders' health risks, resulting in improved risk mitigation, fairer premium calculations, and optimized insurance offerings.

3.2. Individuals/Policyholders:

→ Market Need: Individuals seeking life insurance coverage are often required to undergo complex and time-consuming medical examinations. This process can be intrusive, inconvenient, and may discourage some individuals from obtaining life insurance. A health scoring system provides a more streamlined and non-invasive alternative, allowing individuals to receive personalized life insurance coverage based on their health scores, reducing the need for extensive medical examinations.

→ Customer Need: Individuals desire a simplified and transparent process for obtaining life insurance coverage. A health scoring system caters to this need by providing an accessible platform where individuals can input their health-related information and receive a health score without extensive medical procedures. This empowers individuals to make informed decisions about life insurance and potentially obtain coverage more easily and quickly.

4. TARGET SPECIFICATION AND CHARACTERISATION

4.1. Scalability: The project should be designed and developed with scalability in mind. It should be capable of handling a large volume of data and user requests efficiently. The system should be able to accommodate a growing number of policyholders and seamlessly integrate with insurance companies' existing platforms or APIs.

4.2. Accuracy and Predictiveness: The health scoring system should strive for high accuracy in predicting the health risks of individuals. The machine learning model used for scoring should be trained on a comprehensive dataset and optimized to minimize false positives and false negatives. The system should provide reliable and trustworthy health scores to insurance companies for effective risk assessment.

4.3. User-Friendly Interface: The website or mobile application implementing the health scoring system should have an intuitive and user-friendly interface. It should be easy for individuals to input their health-related information and view their health scores. Insurance companies, brokers, and agents should also have access to a user interface that allows them to understand and interpret the health scores effectively.

4.4. Privacy and Security: Privacy and security are critical considerations in handling sensitive personal health data. The project should adhere to industry-standard practices for data encryption,

secure data storage, and secure data transmission. It should comply with relevant data protection regulations, to ensure the privacy and confidentiality of individuals' health information.

4.5. Integration and Interoperability: The health scoring system should be designed for seamless integration and interoperability with insurance companies' systems and workflows. It should provide APIs or integration mechanisms that allow easy data exchange and communication between the health scoring system and insurers' existing platforms. This ensures smooth adoption and utilization of the system within the insurance industry.

4.6. Performance and Efficiency: The project should prioritize performance and efficiency to provide timely health scores and handle real-time calculations. The machine learning model should be optimized for fast inference, and the system should have minimal response times to provide instantaneous health scores to users. Efficient data processing and storage mechanisms should be implemented to minimize computational resources.

4.6. Adaptability and Updates: The project should be adaptable to changing healthcare and insurance industry trends. It should be designed to incorporate updates and advancements in medical knowledge, evolving risk factors, and changing insurance practices. The system should have mechanisms in place to handle model updates and ensure ongoing relevance and accuracy of the health scoring process.

5. EXTERNAL SEARCH

The source of reference that I used for the motivation of this product is a website called omni calculator (https://www.omnicalculator.com/health), which is an open source that calculates various body measurements such as body fat, body mass index, calories needed, body shape calculator, ideal weight calculator, and so on based on the data entered by the user. They say that this website is used for precise body measurements by medical professionals, students, and patients. In addition, based on the results, they recommend modifications in diet, exercise, and medical exams. Such calculators are quite handy for individuals because they are simple to use and claim to provide very precise results.
This website has inspired me to create a model, iterate it in a mobile app or website, and make it simple to use while also monetising the idea by cooperating with insurance providers.

6. STEPS TO DEVELOP THIS PRODUCT

6.1. Data Collection and Processing:

→ Gather a large dataset containing information on individuals' medical history, pathological test results, lifestyle choices (e.g., smoking, alcohol consumption, addictions), occupation, and environmental factors.

→ Preprocess the collected data to ensure consistency, remove outliers, handle missing values, and transform it into a suitable format for training machine learning models.

6.2. Feature Selection and Engineering:

→ Conduct a comprehensive analysis of the collected data to identify the most relevant features that significantly impact an individual's health and the likelihood of future illnesses.

→ Apply feature engineering techniques to extract meaningful insights from the available data, such as creating new features or aggregating existing ones, to enhance the predictive power of the model.

6.3. Machine Learning Model Development:

→ Select an appropriate machine learning algorithm, such as logistic regression, random forest, or a neural network, to build a predictive model for health scoring.

→ Train the model using the preprocessed dataset, with the target variable representing the probability or severity of potential future illnesses.

→ Utilize appropriate evaluation metrics to assess the performance of the model, such as accuracy, precision, recall, and area under the ROC curve.

6.4. Health Scoring System Implementation:

→ Develop a website or a mobile application where users can input their relevant health-related information.

→ Implement the trained machine learning model into the system to calculate the health score for an individual based on the provided data.

→ Design an intuitive user interface that presents the health score in a clear and understandable manner, allowing insurers to make informed decisions regarding life insurance coverage.

6.5. Integration and Deployment:

→ Integrate the developed health scoring system with insurance companies' existing platforms or establish appropriate APIs for seamless data exchange.

→ Ensure the system's scalability, efficiency, and security to handle a large number of requests and safeguard sensitive personal health data.

→ Deploy the system to a production environment, conducting thorough testing and monitoring to ensure its stability and reliability.

7. APPLICABLE REGULATIONS (GOVERNMENT AND ENVIRONMENT)

7.1. Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011:

→ These rules under the Information Technology Act, 2000, outline security practices and procedures to protect sensitive personal data or information. Compliance with these rules is essential to ensure the security and confidentiality of individuals' health-related data processed by the health scoring system.

7.2. National Digital Health Mission (NDHM):

→ The NDHM is an initiative of the Indian government to establish a digital health ecosystem and promote the use of technology in healthcare. While not directly applicable to insurance companies, the NDHM's guidelines may influence the collection and sharing of health-related data. It is important to stay updated on the NDHM's policies and guidelines to align the health scoring system with the broader digital health initiatives in India.

7.3. Insurance Regulatory and Development Authority of India (IRDAI):

→ The IRDAI is the regulatory body governing the insurance industry in India. It sets guidelines and regulations related to underwriting practices, risk assessment, and data protection for insurance companies. It is essential to comply with the IRDAI's guidelines while developing the health scoring system to ensure adherence to the regulatory framework.

8. APPLICABLE CONSTRAINTS

The training dataset is the most difficult constraint of this project because medical history and an individual's medical records are very sensitive and people are usually not in favour of sharing the data with a commercial industry, and in order to calculate the health score accurately, we need an authentic and large amount of data and cannot simply rely on the public dataset. There are several rules in our country and around the world to protect individuals' privacy, which we also respect, so obtaining this much data would be difficult.

Another constraint is that the score calculation method and outcomes must be evaluated by a pathologist, particularly during the early stages of product development, because we cannot blindly follow our machine learning model; the procedure must be verified by an expert pathologist.
The third constraint is that this score is only a probability of a person becoming sick in the future and is not 100% accurate, so insurance companies should not use it to scare people about their health to make more money. It is a tool to help companies identify customers, not to scare them.

9. BUISNESS OPPORTUNITY

One approach is to offer the health score calculator as a software-as-a-service (SaaS) subscription model. Insurance companies can subscribe to the service and gain access to the platform, allowing them to generate health scores for potential policyholders. The subscription fee can be structured based on factors such as the number of health scores generated, the frequency of usage, or the size of the insurance company. This model ensures a recurring revenue stream for the company while providing insurance companies with a cost-effective solution that enhances their risk assessment and underwriting processes.

Furthermore, partnering with insurance brokers or agents can create opportunities for revenue sharing. The company can establish strategic alliances, allowing brokers and agents to leverage the health score calculator as part of their service offerings. In return, our company can receive a percentage of the premiums generated from policies underwritten using the health score calculator.

10. CONCEPT GENERATION

Combining machine learning models and expert-based scoring in the calculation of a health score within a health scoring system can be achieved through a technical integration approach. The integration process involves leveraging the predictive capabilities of machine learning models and incorporating expert-defined criteria to refine the results. Here's a technical overview of the combination:

10.1. Machine Learning Models:

- → Develop machine learning models, such as classification or regression algorithms, to analyze large datasets and predict health risk probabilities for individuals. These models can be trained using historical health records, lifestyle information, and environmental factors as input features. The models learn patterns and relationships within the data to generate quantitative health risk predictions.

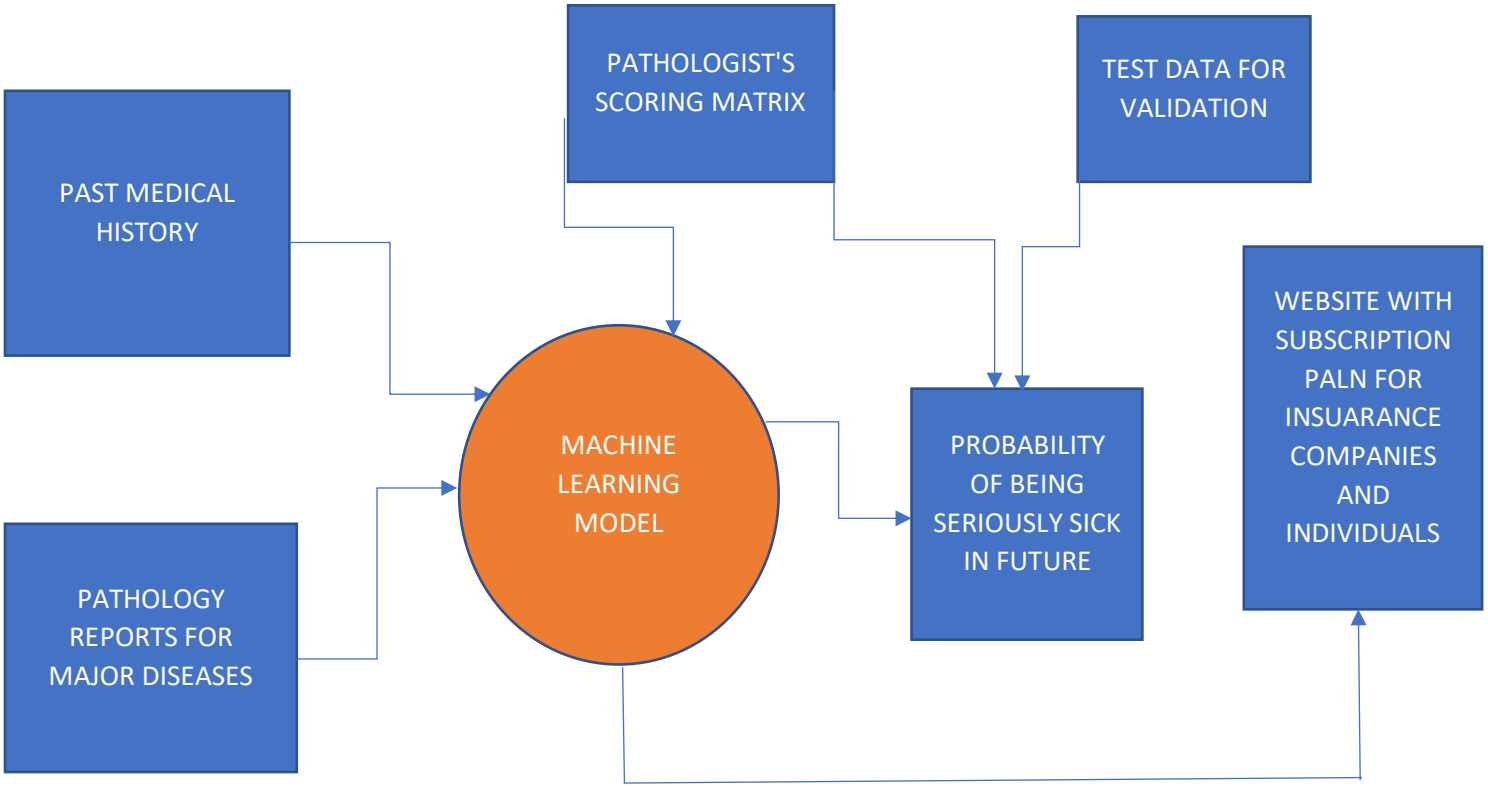
10.2. Expert-Based Scoring:

- → Collaborate with medical experts to define scoring criteria based on their domain knowledge. Identify key health parameters, risk factors, and thresholds that hold significance in determining health status. Assign scores or weights to each criterion according to their relative importance and impact. These criteria capture qualitative insights and expert judgment, complementing the data-driven approach of the machine learning models.

10.3. Combination Technique:

→ Integrate the machine learning predictions and expert-based scoring using a combination technique. One approach is to adjust the machine learning predictions based on expert feedback or weights assigned to each criterion. This can be achieved through techniques like feature scaling or incorporating expert-defined thresholds into the prediction output. Another technique is to include expert-derived factors as additional inputs to the machine learning models, allowing the models to learn from both data-driven and expert-defined features

11. FINAL PRODUCT PROTOTYPE



12. TEAM REQUIRED TO DEVELOP THIS PRODUCT

12.1. Data Scientists:

→ Data scientists play a crucial role in analyzing and modeling large health-related datasets. They should possess skills in data preprocessing, feature engineering, and machine learning techniques. Expertise in developing predictive models and selecting appropriate algorithms for health risk assessment is essential.

12.2. Medical Professionals:

→ Medical professionals, such as doctors or healthcare practitioners, bring domain knowledge and expertise in understanding health-related factors and risk indicators. Their insights are valuable for defining scoring criteria, thresholds, and weightings that align with medical standards and practices.

12.3. Software Engineers:

→ Software engineers are responsible for implementing the infrastructure and architecture required for the health scoring system. They should have expertise in programming

languages, database management, and cloud computing. Skills in building scalable and secure systems are essential for the successful implementation and deployment of the product.

12.3. User Experience (UX) Designers:

→ UX designers are responsible for creating a user-friendly and intuitive interface for the health scoring system. They should have a deep understanding of user needs, information architecture, and visual design principles. Their expertise ensures a seamless and engaging user experience that enables easy interpretation of health scores and insights.

12.4. Data Privacy and Security Specialists:

→ Data privacy and security specialists are essential to ensure compliance with regulations and safeguard individuals' sensitive health information. They should have expertise in data protection laws, encryption techniques, access controls, and secure data storage and transmission.

12.5. Project Manager:

→ A project manager oversees the entire development process, ensuring effective coordination among team members, setting project goals and timelines, and managing resources and stakeholder expectations. They should possess strong communication, leadership, and project management skills.

13. CONCLUSION

The product ideation project for the health scoring system presents a groundbreaking approach that merges the strengths of machine learning models and expert-based scoring to revolutionize risk assessment in the insurance industry. By leveraging the capabilities of machine learning, the system can analyze vast and complex datasets, extracting patterns and correlations that contribute to accurate health risk predictions. This data-driven approach provides objective and quantifiable health scores based on an individual's medical records, lifestyle factors, and environmental influences. Furthermore, the integration of expert-based scoring brings valuable qualitative insights and domain expertise, enhancing the system's precision and contextual understanding.

The prototype of the health scoring system showcases its potential to transform risk assessment and underwriting practices in insurance. It empowers insurance companies with a sophisticated tool to make informed decisions and offer tailored coverage to policyholders. The combination of machine learning models and expert-based scoring ensures a comprehensive evaluation of health risks, enabling insurers to better understand an individual's health profile and anticipate potential risks. By providing a holistic view of health conditions and leveraging both quantitative and qualitative assessments, the system contributes to more accurate risk evaluation and enables personalized insurance solutions. As the system evolves through continuous iteration and refinement, guided by user feedback and emerging research, it holds the promise of becoming a robust and impactful solution that shapes the future of risk management and promotes healthier lives for individuals across the insurance landscape.

13. REFERENCES

Customer heath score. Link - https://www.pendo.io/glossary/customer-health-score/

How Your Credit Score Impact Your Financial Future. Link - https://www.finra.org/investors/personal-finance/how-your-credit-score-impacts-your-financial-future#Infographic-ThinkingMoney

Sun, H., Depraetere, K., Meesseman, L., Cabanillas Silva, P., Szymanowsky, R., Fliegenschmidt, J., Hulde, N., von Dossow, V., Vanbiervliet, M., De Baerdemaeker, J., Roccaro-Waldmeyer, D. M., Stieg, J., Domínguez Hidalgo, M., & Dahlweid, F. M. (2022). Machine Learning-Based Prediction Models for Different Clinical Risks in Different Hospitals: Evaluation of Live Performance. *Journal of medical Internet research*, *24*(6), e34295.

https://doi.org/10.2196/34295

Mental Health Prediction Using Machine Learning. Link- https://www.analyticsvidhya.com/blog/2022/06/mental-health-prediction-using-machine-learning/

Financial analysis and health score for buying insurance? This company CIBIL like solution.

Link - https://www.moneycontrol.com/news/business/economy/facial-analysis-and-health-score-for-buying-insurance-this-company-is-offering-a-cibil-like-solution-6646661.html

Omni calculator. Link - https://www.omnicalculator.com/

List of life insurers, Insurance Regulatory and Development Authority of India, Govt, of India

Link - https://irdai.gov.in/list-of-life-insurers1