Enhancing Insurance Industry with Machine Learning for Expense Prediction

Team Members

Ritesh Kate

Abhishek Kumar

Piyush Singh

Chirag Jain

Alen Sadique

1 Prototype Selection

Abstract

We have chosen the prototype for our financial modelling based on the specific characteristics of our market and product/service. The selection of this prototype aligns with the observed growth pattern in our target market. For a linear market, where sales grow steadily over time, our financial model provides a straightforward representation, accounting for pricing, sales, and costs. On the other hand, in cases of exponential market growth, the chosen prototype accommodates the expected rapid expansion by incorporating an exponential equation for sales. This selection ensures that our financial modelling approach is adaptable to both linear and exponential scenarios, allowing us to accurately project total profit under varying market conditions. Moreover, this prototype offers clarity and simplicity in its representation, making it accessible for stakeholders to understand and work with, facilitating better decision-making and planning for our business.

2 Problem Statement:

Insurance companies, especially when it comes to life insurance, struggle to figure out how healthy someone is and how likely they are to get seriously sick in the future. The way they currently do this isn't very accurate and often involves time-consuming and intrusive medical check-ups for people applying for insurance. We need to find a better way to do this. The problem we're trying to solve is to create a smart computer system that can use lots of health-related information to give a clear "health score" for individuals. This score would help insurance companies make fair decisions about who to insure and how much to charge them. The challenge is to make this system easy to use, reliable, and fair for both insurance companies and people who want life insurance.

3 Market/Customer/Business Need Assessment:

For Insurance Companies:

Why Insurance Companies Need It: Insurance companies want a better way to understand the health risks of people who want insurance. Right now, they use old methods that aren't very good at this.

How It Helps Them: This new system can help insurance companies make smarter decisions. They can offer insurance to the right people at the right price, which is good for their business.

What They Want: Insurance companies need a system that's easy to use and gives them accurate information about a person's health. They also want it to be quick and not cost too much.

For Individuals/Policyholders:

Why Individuals Need It: People who want life insurance often have to go through lots of medical tests, which can be bothersome and take a long time.

How It Helps Them: With this system, they might not need to do all those tests. It can make getting life insurance easier and faster.

What They Want: Individuals want a simple and clear way to show that they're healthy. They want to get insurance without too much hassle.

For the Business:

Why the Business Needs It: Our business wants to create something that helps insurance companies and people. It's a win-win. Insurance companies can make better choices, and people can get insurance more easily.

How It Helps the Business: If we build a good system, insurance companies will pay to use it. We can also work with insurance agents who help people get insurance, and they can use our system too.

What We Want: We want to create a system that's really good at predicting health risks. We want to make money by offering it to insurance companies and agents.

4 . Target Specifications and Characterization

Scalability:

Handles extensive data and numerous users without performance degradation, ensuring smooth growth. This means the system can accommodate a growing number of users and data inputs without slowing down, providing a responsive experience as it scales.

Accuracy and Predictiveness:

Achieves precise predictions of future health risks, facilitating insurance decisions and error prevention. The system's accuracy ensures that insurance companies can rely on its predictions, reducing the chances of making costly mistakes.

User-Friendly Interface:

Offers an intuitive and straightforward user experience for all, ensuring ease of use by individuals and insurance companies. A user-friendly

interface means that both individuals and insurance professionals can navigate the system without extensive training, enhancing accessibility.

Privacy and Security:

Ensures the confidentiality and protection of health information, establishing trust and data security. Robust privacy and security measures safeguard sensitive health data, instilling trust among users and complying with data protection regulations.

Integration and Interoperability:

Seamlessly interacts with existing insurance systems, streamlining processes. Integration capabilities allow for the smooth exchange of data with insurance company systems, reducing workflow disruptions and enhancing efficiency.

Performance and Efficiency:

Operates swiftly and efficiently, optimising resource utilisation. High-performance and resource-efficient operation ensure that the system responds quickly and doesn't strain computational resources.

Adaptability and Updates:

Evolves and improves over time to remain relevant and valuable as industry practices evolve. The system's adaptability enables it to stay up-to-date with changing healthcare and insurance industry trends, maintaining its effectiveness and usefulness**.**

5 Prototype Development

The project at hand centres around the ambitious goal of predicting health insurance expenses, a critical endeavour within the ever-evolving healthcare sector. Employing advanced machine learning techniques and leveraging a diverse dataset brimming with pertinent customer attributes, this project seeks to revolutionise the way health insurance expenses are managed and forecasted. By harnessing the power of predictive modelling, the aim is to provide insurance providers, policymakers, and healthcare professionals with invaluable insights into cost projections. This, in turn, can empower them to make informed decisions, optimise resource allocation, and enhance the overall efficiency of healthcare delivery.

In a world where healthcare costs continue to rise, accurate expense predictions hold immense potential for reducing financial burdens on both insurance companies and individuals. This project not only addresses a pressing industry need but also signifies a significant step towards a more cost-effective and sustainable healthcare ecosystem.

Dataset Description

The dataset under consideration includes the following columns:

Age: Age of the insured individual.

Sex: Gender of the insured individual.

BMI: Body Mass Index of the insured individual.

Children: Number of children/dependents covered by the insurance.

Smoker: Smoking status of the insured individual (smoker or non-smoker).

Region: Region in which the insured individual resides.

Expenses: Insurance expenses incurred by the insured individual.

Correlation Analysis

To assess the feasibility of building an ML model for predicting insurance expenses, we conducted a correlation analysis between the "Expenses" column (our target variable) and the other independent variables:

In [4]:

```
import numpy as np
import pandas as pd
data = pd.read_csv("/content/insurance (1).csv")
data.head()
```

Out[4]:

	age	sex	bmi	children	smoker	region	expenses
0	19	female	27.9	0	yes	southwest	16884.92
1	18	male	33.8	1	no	southeast	1725.55
2	28	male	33.0	3	no	southeast	4449.46
3	33	male	22.7	0	no	northwest	21984.47
4	32	male	28.9	0	no	northwest	3866.86

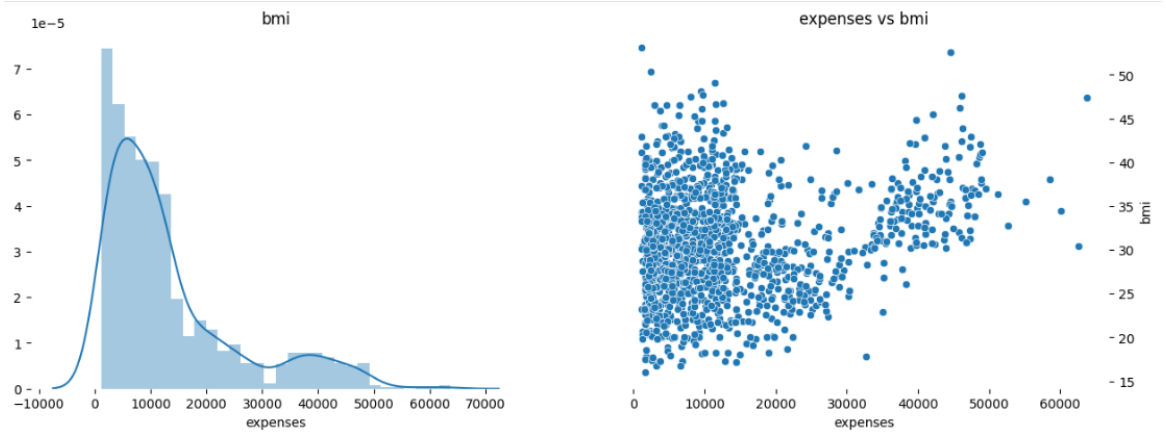
In [5]:

```
data.isnull().sum()
```

Out[5]:

```
age         0
sex         0
bmi         0
children    0
smoker      0
region      0
expenses    0
dtype: int64
```

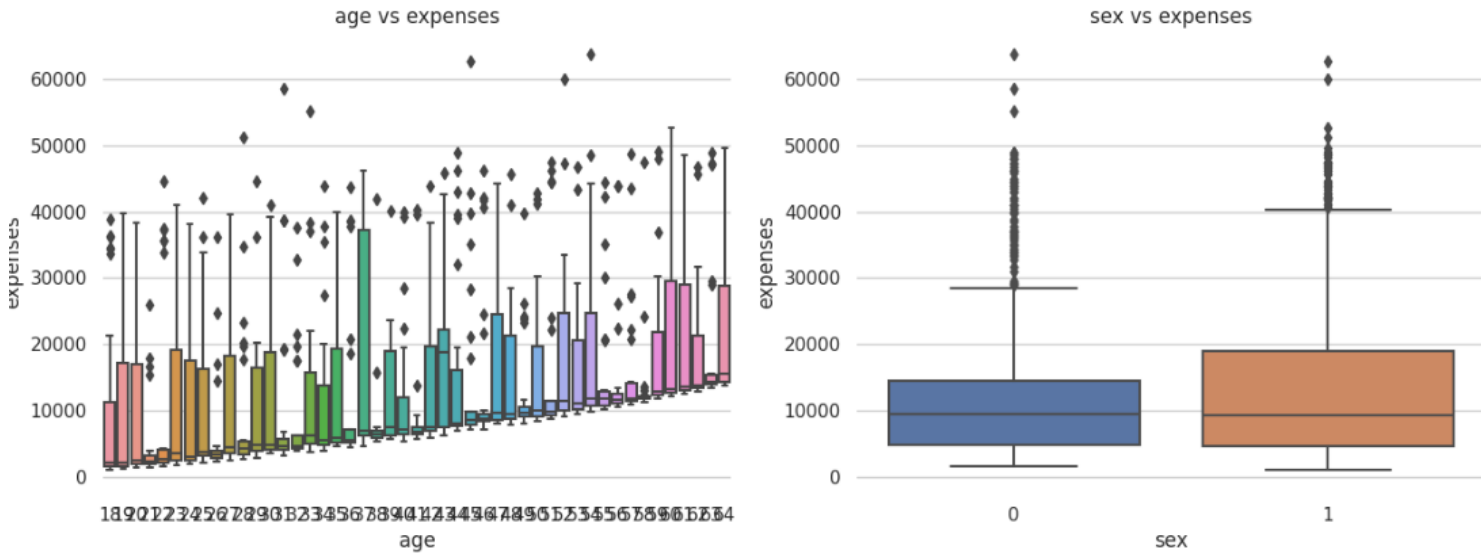
```
f, axes = plt.subplots(1, 2,figsize=(15,5))
sns.distplot(df['expenses'], ax=axes[0])
sns.scatterplot(x='expenses',y='bmi', data=df, ax=axes[1])
sns.despine(bottom=True, left=True)
axes[0].set(xlabel='expenses', ylabel='', title='bmi')
axes[1].set(xlabel='expenses', ylabel='bmi', title='expenses vs bmi')
axes[1].yaxis.set_label_position("right")
axes[1].yaxis.tick_right()
```




```
sns.set(style="whitegrid", font_scale=1)

f, axes = plt.subplots(1, 2,figsize=(15,5))
sns.boxplot(x=df['age'],y=df['expenses'], ax=axes[0])
sns.boxplot(x=df['sex'],y=df['expenses'], ax=axes[1])
sns.despine(bottom=True, left=True)
axes[0].set(xlabel='age', ylabel='expenses', title='age vs expenses')
axes[1].set(xlabel='sex', ylabel='expenses', title='sex vs expenses')
```

```
[Text(0.5, 0, 'sex'),
 Text(0, 0.5, 'expenses'),
 Text(0.5, 1.0, 'sex vs expenses')]
```



Industrial Significance

Predicting health insurance expenses addresses several critical problems and holds significant industrial significance:

1. Cost Management for Insurers: Insurance companies often struggle with accurately forecasting healthcare costs, leading to financial challenges and the risk of underpricing or overpricing policies. Predictive modelling helps insurers better estimate future expenses, ensuring they can set premiums at appropriate levels to remain financially viable.

2. Premium Stability: For policyholders, stable and predictable premium rates are crucial. Accurate expense predictions enable insurers to offer more stable premiums, reducing the financial burden on individuals and businesses purchasing health insurance.

3. Resource Allocation: Healthcare providers, including hospitals and clinics, can use expense predictions to allocate resources effectively. They can anticipate patient volumes and the types of services required, optimising staff schedules, equipment usage, and inventory management.

4. Risk Mitigation: Health insurance predictions allow insurers to identify high-risk individuals or groups, enabling them to implement targeted interventions and preventive measures. This reduces the overall risk pool and can lead to lower costs for all policyholders.

5. Policy Development and Regulation: Governments and regulatory bodies can use expense predictions to inform policy development and healthcare regulations. This ensures the industry remains sustainable and accessible while protecting consumer interests.

6. Data-Driven Decision-Making: The industrial significance lies in the transformation of the healthcare and insurance sectors into data-driven industries. Predictive modelling encourages evidence-based decision-making, fostering efficiency and innovation.

7. Improved Patient Outcomes: By accurately estimating future healthcare expenses, insurers and providers can invest in programs that improve patient outcomes, preventive care, and early intervention, ultimately reducing costs and enhancing the quality of care.

8. Market Competitiveness: Companies that can offer more accurate and competitive insurance products are likely to gain market share. Accurate expense predictions give insurers a competitive edge in a crowded marketplace.

9. Long-Term Sustainability: The healthcare industry faces increasing financial pressures due to factors like an ageing population and rising medical costs. Predictive modelling contributes to the long-term sustainability of the industry by helping stakeholders adapt to these challenges proactively.

10. Research and Development: Expense predictions can also support pharmaceutical and medical device companies in targeting their research and development efforts more effectively. This leads to the development of cost-effective treatments and therapies.

In summary, predicting health insurance expenses is a solution to the complex challenge of managing healthcare costs, benefiting insurance companies, healthcare providers, policymakers, and patients alike. It enhances financial stability, supports better decision-making, and contributes to the overall sustainability and efficiency of the healthcare and insurance industries.

6. BENCHMARK

Data Collection Milestone:

Collect a diverse dataset of health-related information from various sources, including medical history, lifestyle, and environmental factors, for a comprehensive analysis.

Machine Learning Model Development:

Develop and train a machine learning model to predict health risks with a minimum accuracy of 85% based on historical data.

System Prototype:

Create a working prototype of the health scoring system in the form of a functional website or mobile application.

Privacy and Security Compliance:

Ensure the system complies with privacy and security regulations, obtaining the necessary certifications or approvals for handling sensitive health data.

User Testing and Feedback:

Conduct user testing with a minimum of 100 users to gather valuable feedback for system improvements.

Integration with Insurance Companies:

Successfully integrate the system with at least three insurance companies' existing platforms, enabling seamless data exchange.

Scalability Test:

Test the system's ability to handle a growing user base, ensuring it can accommodate a load of 10,000 simultaneous users without performance degradation.

Performance Optimization:

Optimise system performance to achieve an average response time of under 2 seconds for health score calculations.

Adaptability and Updates:

Implement a mechanism for regular updates and improvements, aiming to release at least three major updates in the first year to adapt to changing health and insurance trends.

Business Adoption:

Secure contracts with at least five insurance companies and 20 insurance agents within the first year of operation, demonstrating successful adoption by the insurance industry.

7. BUSINESS MODEL

1. **Value Proposition:**

Provide insurance companies with a data-driven and objective health scoring system to assess potential policyholders' health risks accurately. Offer individuals a streamlined and non-invasive alternative to traditional medical examinations when applying for life insurance.

2. **Customer Segments:**

Insurance Companies: Target insurance companies, both established and emerging, looking to enhance their underwriting practices and offer personalised life insurance coverage. Insurance Brokers and Agents: Partner with insurance brokers and agents to extend the reach of the health scoring system.

3. **Channels:**

Direct Sales Team: Use a direct sales team to approach insurance companies and demonstrate the value of the health scoring system. Online Marketing: Utilise online marketing channels such as social media, content marketing, and search engine optimization to reach a broader audience.Partnerships: Collaborate with industry associations, insurance broker networks, and healthcare organisations to promote the product.

4. **Customer Relationships:**

Provide customer support to assist insurance companies and brokers in using the system effectively.Gather feedback from users and continuously improve the product.

5. **Key Activities:**

Data Acquisition and Processing: Continuously collect and preprocess health-related data to ensure accuracy and relevance Model Development: Develop and fine-tune machine learning models for health scoring.System Implementation: Build and maintain the website or mobile

application for generating health scores.Compliance and Security: Ensure compliance with data protection regulations and maintain robust data security.

6. **Key Resources:**

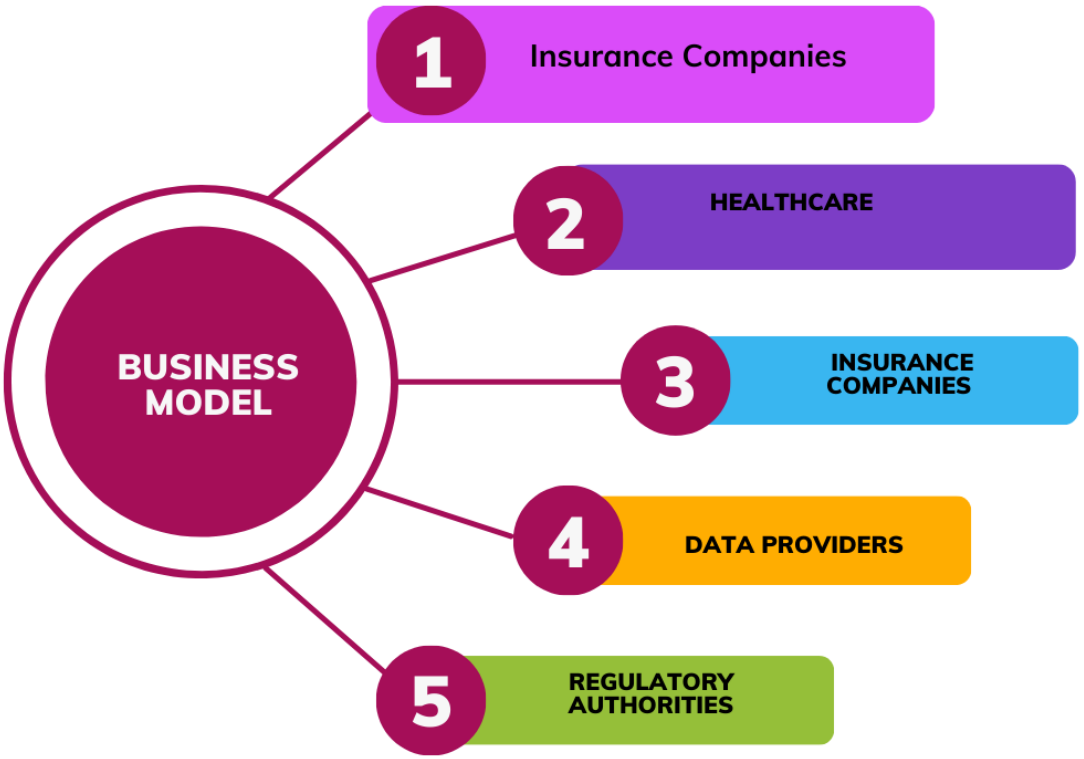
Data Sources: Secure access to a diverse and extensive dataset for training and validation. Skilled Workforce: Assemble a team of data scientists, medical professionals, software engineers, UX designers, and security specialistsTechnology Infrastructure: Invest in the necessary technology and infrastructure to support the system.

7. **Partnerships:**

Partner with data providers, healthcare institutions, and insurance industry associations to access data and reach potential customers.

8. **Cost Structure:**

Data Acquisition: Invest in acquiring and maintaining the dataset.Development and Maintenance: Allocate resources for system development and ongoing maintenance Security and Compliance: Budget for data privacy and security measures.Marketing and Sales: Allocate funds for marketing and sales efforts.



8 Financial Model:

1. Revenue Streams:

Subscription Model: Offer a subscription-based pricing model for insurance companies. They pay a monthly or annual fee based on the number of health scores generated or the size of their policyholder base.

Revenue Sharing: Partner with insurance brokers or agents to generate revenue through a percentage of the premiums generated from policies underwritten using the health score calculator.

2. Cost Structure:

Data Acquisition: Invest in acquiring large health-related datasets, which may include the cost of obtaining medical records, pathological test results, and lifestyle data.

Development and Maintenance: Expenses related to the development, maintenance, and continuous improvement of the health scoring system, including salaries for data scientists, software engineers, and UX designers.

Security and Compliance: Costs associated with ensuring data privacy and security compliance, including hiring data privacy and security specialists.

Marketing and Sales: Expenses for marketing the product to insurance companies and brokers, as well as sales efforts to acquire new subscribers.

3. Key Performance Indicators (KPIs):

Subscriber Growth Rate: Monitor the growth in the number of insurance companies subscribing to the service.

Churn Rate: Track the rate at which insurance companies discontinue their subscriptions.

Average Revenue Per User (ARPU): Calculate the average revenue generated from each subscribing insurance company.

Customer Acquisition Cost (CAC): Measure the cost of acquiring a new insurance company subscriber.

Lifetime Value (LTV): Estimate the total revenue generated from a subscribing insurance company throughout their partnership.

Let's take an example and calculate it:

1. Monthly Subscription Fee per Insurance Company (F): $5

2. Number of Subscribing Insurance Companies (N):

Month 1: 10

Monthly Churn Rate (percentage of companies discontinuing subscription): 5%

Month	New Subscriptions	Churned Subscriptions	Total Subscriptions
1	10	0	10
2	5	1 (5% of 10)	14
3	5	1 (5% of 14)	18
4	5	1 (5% of 18)	22
5	0	1 (5% of 22)	21
6	5	0	26
7	0	0	26
8	0	0	26
9	5	0	31
10	0	0	31
11	0	0	31
12	5	0	36

3. Monthly Revenue (R):Calculate the monthly revenue for each month using the formula:

R = F * N

Month	Monthly Revenue (R)
1	$50
2	$70
3	$90
4	$110
5	$105
6	$130
7	$130
8	$130
9	$155
10	$155
11	$155
12	$180

4.Annual Revenue (AR): Calculate the annual revenue by multiplying the monthly revenue by 12:

AR = R * 12

Month	Monthly Revenue (R)	Annual Revenue (AR)
1	$50	$600
2	$70	$840
3	$90	$1,080
4	$110	$1,320
5	$105	$1,260
6	$130	$1,560
7	$130	$1,560
8	$130	$1,560
9	$155	$1,860
10	$155	$1,860
11	$155	$1,860
12	$180	$2,160

9 CONCLUSION

The development of the Health Score Calculator for Insurance Companies represents a significant opportunity to revolutionise the insurance industry's approach to underwriting and risk assessment. By harnessing the power of data, machine learning, and user-friendly interfaces, this project addresses critical market and customer needs.Insurance companies stand to benefit from accurate health risk predictions, streamlined underwriting processes, and improved risk management. This innovative solution enhances their ability to offer tailored life insurance coverage, setting them apart in a competitive market.Individuals seeking life insurance also gain from a simplified, non-invasive approach that eliminates the need for extensive medical examinations.