Heart Attack Risk Prediction: A Comparative Analysis of Machine Learning Models, Performance Metrics, and Cost-Benefit Impact in Healthcare Insurance

Masters Business Analytics September 2024

Author **Rahul Pagar**

Current Scenario

An insurance company wants to reduce the financial losses caused by heart attack claims. Currently, they rely on traditional methods to assess risk, which may not accurately predict which customers are at higher risk.

Traditional screening misses 10% of High-risk individuals.

Problem statement:

The company aims to use machine learning to predict which customers have a higher risk of heart attacks. Reduce the number of claims and associated costs by offering early interventions.

Model Selection & Evaluation Metrics

Model evaluated	**Logistic Regression:** A simple, interpretable model suitable for predicting heart attack risk when there is a linear relationship between features and the target.
	XGBoost : A high-performance, robust model that can capture complex, non-linear relationships and handle large datasets with good performance in heart attack risk prediction **(Final Selected Model)**
	KNN: A flexible model that classifies heart attack risk based on proximity to similar individuals.
Evaluation Metrics Considered	**Recall (Sensitivity):** Recall is crucial because it emphasizes identifying as many positive cases (true heart attack risks) as possible. Missing a positive case (False Negative) can have serious consequences, so maximizing recall helps ensure high sensitivity in detecting individuals at risk.

Model	Recall	Precision
Logistic Regression	0.65	0.33
XG Boosting	0.72	0.34
KNN	0.54	0.33

Confusion Matrix

XGBoost– Results on Test Data
Key Observations:
True Positives (TP = 441): Correctly identified high risk individuals.

False Negatives (FN = 170): Missed high risk individual.

False Positives (FP = 851): Incorrectly identified high risk a high-risk individual.

True Negative (FP = 291): Correctly identified low risk individuals.

Why Recall Matters?

False Negatives (Missed Cases) are Risky: Undiagnosed obese clients lead to late-stage health issues, increasing insurance claims.

False Positives (Unnecessary Interventions) are Less Harmful: The cost of additional screening is lower than the cost of missing a high-risk individual.

	Predicted Heart Attack Risk	**Predicted No Heart Attack Risk**
Actual Heart Attack Risk	TP = 441	FN = 170 (missed cases)
Actual No Heart Attack Risk	FP = 851 (false alarms)	TN = 291

Significant predictors for heart risk selected by XG boost	
Sleep Hours Per Day	0.701391
Diabetes	0.103895
Systolic	0.068945
Physical Activity Days Per Week	0.066571
Diastolic	0.059197

Cost-Benefit Analysis

Current Approach (Without Model)		
	Missed Heart attack risk clients	876 (10% of actual cases)
	Cost per heart attack claim	€5,0000
	Expected claims payout	€43,800,000
	Cost per policyholder for cardiac tests (8764 x €500)	€43,820,000
	Total Cost	€ 87,620,000

With Neural Network Model		
	Total Intervention cost (TP+FP) x €500	€6,46,000
	Reduction in claims payout (assuming 25% of flagged individuals take preventive measures and avoid claims) 50% of 441 TP = 110 saved from claims Savings = 110 x € 50,000	€ 5,500,000
	Net claims payout with model €43,800,000 – 5,500,000	€ 38,300,000
	Total Savings in Claims	€ 5.5 million

	Without model	With model
Heart Attack claims Payout	€43,800,000	€ 38,300,000
Unnecessary test	€43,820,000	€6,46,000
ML Implementation	-	€3500000
Total Cost	€ 87,620,000	€36,946,000

Total saving per year after implementation of ML model €50.67M

ROI after implementation of ML model 751%

Conclusion

By implementing the ML-based heart attack risk prediction system, the insurance company can save approximately **€ 50.67 million per year** despite additional implementation costs. The system significantly improves early detection, reducing undetected cases and claim payouts. However, high false positive rates result in unnecessary testing costs, which should be minimized by improving model precision.

Future Scope : Increasing **recall** ensures fewer missed heart attack cases, reducing severe claim costs. Enhancing **precision** will lower false positives and prevent unnecessary medical expenses.
Overall, this system provides a **high ROI** while improving policyholders' health outcomes and reducing insurance liabilities