IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

Insurance Claim Prediction

Varshini H S[1] and Dr. Chitra K[2]
Student MCA, IVth Semester[1]
Associate Professor, Department of MCA[2]
Dayananda Sagar Academy of Technology and Management, Udaypura, Bangalore, Karnataka, India
varshinihs455@gmail.com and chitra-mca@dsatm.edu.in

Abstract: *The Health Insurance Claim Prediction system forecasts insurance claims based on health metrics using Flask and MySQL. It features secure client registration, login, and data management. Bulk information can be added through CSV uploads, maintaining an up-to-date database essential for accurate predictions. The dashboard provides key metrics like user demographics for quick insights. Predictive modeling uses SVM and KNN algorithms trained on historical data to forecast insurance claims, offering real-time feedback based on inputs like age, sex, and BMI. Data preprocessing ensures clean and analyzable data. This system helps insurance companies assess risk, tailor plans, and improve customer service*

Keywords: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Flask, Prediction, Insurance

I. INTRODUCTION

The Wellbeing Protections Claim Expectation framework may be a modern stage outlined to function on a nearby have, centering on foreseeing protections claim probabilities based on different wellbeing measurements. It leverages Jar, a small scale web system for Python, to coordinated consistently with a MySQL database, guaranteeing effective information administration and exact forecasts. The framework offers a extend of user-friendly highlights, counting enlistment, login, and contact shapes, encouraging secure information capacity and administration. It bolsters bulk information addition through CSV record transfers, empowering fast integration of expansive datasets. Clients can clear datasets as required, keeping up an up-to- date database fundamental for precise forecasts.

The framework highlights a dashboard that outwardly shows key measurements, such as the number of male and female clients, giving prompt bits of knowledge into statistic dissemination. This helps clients in analyzing patterns and designs successfully. At the heart of the framework lies prescient modeling, utilizing machine learning calculations like Bolster Vector Machine (SVM) and K-Nearest Neighbors (KNN) to figure protections claims. These models, prepared on chronicled information, recognize designs and connections between wellbeing parameters and claim probabilities. By contributing measurements like age, sex, and BMI, clients get real-time expectations approximately their claim probability, enabling educated wellbeing and protections choices.

Information preprocessing is significant, including reading, cleaning, and embeddings CSV records into the database to guarantee the information is appropriate for investigation. Verifiable information is part into preparing and testing sets to prepare and assess the SVM and KNN models, guaranteeing vigorous expectations. Unused information prompts the framework to inquiry the database for significant authentic records, improving result unwavering quality.

The application underpins Angular-based dashboards and information administration, permitting consistent front-end integration and energetic client interfacing. Its adaptable engineering can be custom fitted to meet different client and organizational needs. For protections companies, the framework gives critical benefits by analyzing client wellbeing information to survey hazard and foresee claims more precisely. This encourages educated decision-making, personalized protections plans, and moved forward client benefit. Policyholders pick up experiences into their wellbeing dangers, advancing proactive wellbeing administration and possibly lessening premiums. The Wellbeing Protections Claim Forecast framework may be a profitable device for improving decision-making in wellbeing protections, profiting both guarantees and policyholders.

DOI: 10.48175/IJARSCT-22541

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

II. LITERATURE SURVEY

Abdi Negara Guci et al. [1] emphasize the crucial role of the health insurance industry in mitigating financial risks. They propose combining Support Vector Machine (SVM) with the Bat Algorithm (BA) to enhance prediction accuracy and efficiency. Their study demonstrates that the SVM-BA model outperforms traditional SVM and SVM-PSO models, reducing computing time and avoiding overfitting.

Akashdeep Bhardwaj's [2] study focuses on predicting annual medical claim expenses for BSP Life (Fiji) Ltd. using artificial neural network (ANN) models. The research demonstrates that recurrent neural networks (RNNs) outperform feedforward networks in accuracy and computational efficiency. RNNs reduced mean absolute percentage error (MAPE) by 11.5%, achieving 93.58% accuracy, compared to 87.85% for feedforward networks.

Deandra Aulia Rusdah and Hendri Murfi's [3] research examines predicting insurance risk levels using XGBoost, which can inherently handle missing values through its sparsity- aware algorithm. The study compares scenarios with and without imputation preprocessing, demonstrating that XGBoost maintains high accuracy even without preprocessing. The approach addresses class imbalance using oversampling techniques, highlighting XGBoost's robustness and effectiveness

Endalew Alamir et al. [4] addressed the challenges of classifying and predicting motor insurance claim statuses by developing a robust machine learning model. They used data preparation techniques and K-Fold cross-validation for reliability. Random Forest (RF) and Multi-

Class Support Vector Machine (SVM) were applied, achieving high prediction accuracy of 98.36% and 98.17%, respectively, with RF slightly outperforming SVM.

Keshav Kaushik et al. [5] examined how AI and ML are revolutionizing health insurance by enhancing prediction accuracy and streamlining processes. Their study highlighted an ANN- based regression model predicting premiums with 92.72% accuracy. The model efficiently handles large data volumes, reducing costs and time, and improving overall insurance processes.

Muhammad Arief Fauzan and Hendri Murfi's [6] study highlights the challenges of managing frequent and severe auto insurance claims with missing data. They propose using XGBoost, an ensemble learning method, for accurate predictions. XGBoost outperformed other methods like AdaBoost and Random Forest in normalized Gini accuracy, demonstrating its effectiveness in handling large datasets.

Nidhi Bhardwaj and Rishabh Anand's [7] thesis explores predicting insurance amounts using Multiple Linear Regression, Decision Tree Regression, and Gradient Boosting Decision Tree Regression. Their study found Gradient Boosting Regression to be the most accurate and efficient model. Age and smoking status were identified as the most influential factors in insurance predictions.

Ranjodh Singh et al. [8] propose an automated car insurance claim system using advanced image analysis techniques. By employing models like Mask R-CNN and PANet, along with a VGG16 network, the system efficiently identifies and assesses car damage. It offers significant improvements in claim processing speed and accuracy, enhancing overall insurance operations.

Syaiful Anam et al. [9] explore enhancing health insurance claim predictions by combining Support Vector Machine (SVM) with Particle Swarm Optimization (PSO). The hybrid model improves accuracy significantly compared to standard SVM, though it requires more computation time. PSO helps in finding global optima, thus enhancing predictive performance.

III. METHODOLOGY

3.1 Existing Method

Existing strategies for anticipating protections claims incorporate different machine learning methods outlined to handle the complexities of claim information. Conventional models like Naïve Bayes, Choice Trees, and Counterfeit Neural Systems are commonly utilized. Among these, Bolster Vector Machines (SVM) are outstanding for their adequacy but frequently require ideal parameter tuning through trial and mistake, which can be wasteful. Progressed strategies like Molecule Swarm Optimization (PSO) upgrade SVM by moving forward parameter determination and accomplishing way better expectation exactness. Furthermore, inventive approaches such as robotized image-based harm evaluation and outfit learning strategies have risen, advertising more exact and productive claim preparing.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

3.2 Proposed Method

The proposed method for insurance claim prediction integrates advanced machine learning techniques within a Flask-based web application to enhance claim assessment processes. The system begins by gathering and cleaning historical insurance data stored in a MySQL database. Users interact with the web interface to upload datasets, manage user registrations, and retrieve predictions.

Central to the strategy is the utilize of machine learning calculations, counting Bolster Vector Machine (SVM) and K-Nearest Neighbors (KNN). These calculations are utilized to analyze and foresee protections claims based on different input highlights such as age, sex, BMI, and protections history. The approach includes part the dataset into preparing and testing subsets to assess demonstrate execution. SVM, known for its strength in dealing with high-dimensional information, and KNN, which classifies based on highlight similitude, are both utilized to infer forecasts.

The application provides a seamless experience by allowing users to visualize data, clear or delete records, and receive predictions through an intuitive interface. By automating the prediction process and integrating it with a user-friendly platform, the system aims to streamline claim assessments, reduce processing time, and improve the accuracy of insurance claim predictions.

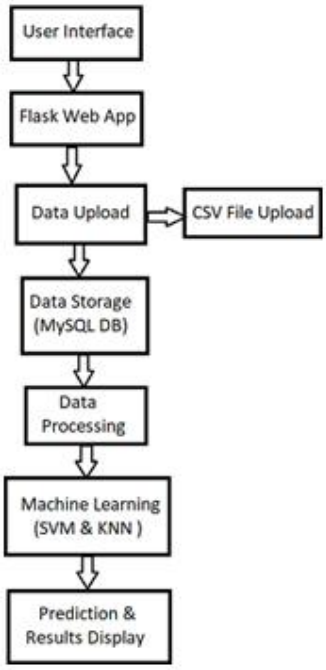

Fig 3.2.1 Block Diagram

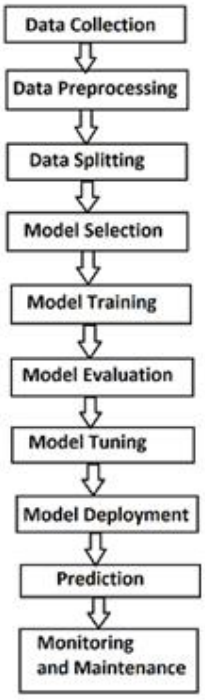

Fig 3.2.2 Prediction Diagram

IV. RESULTS AND DISCUSSIONS

The protections claim expectation framework, created employing a Carafe application, leverages machine learning strategies to assess the likelihood of protections claims based on different individual traits such as age, sex, BMI, smoking propensities, and other significant variables. The workflow starts with the extraction of information from CSV records, which are at that point stacked into a MySQL database for encourage investigation.

The center of the expectation includes two machine learning models:

Back Vector Machine (SVM) and K-Nearest Neighbors (KNN). The SVM show utilizes a direct part to make a choice boundary that ideally isolates diverse classes inside the dataset. Usually especially compelling when the information can be partitioned into particular bunches through a straight line or hyperplane. Then again, the KNN demonstrate classifies unused information focuses by assessing the foremost common lesson among its k-nearest neighbors, with k

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

set to 7 in this case. This strategy benefits from straightforwardness and the capacity to adjust to the neighborhood structure of the information, in spite of the fact that it can be delicate to the choice of k and the highlight scaling.

The models' execution is measured by their precision, determined from comparing their forecasts on test information against genuine results. The precision scores gotten reflect the effectiveness of each demonstrate in anticipating protections claims. Whereas both SVM and KNN illustrate solid prescient capabilities, the precision and unwavering quality of the forecasts are affected by the quality of the information, the choice of highlights, and preprocessing procedures. To improve the framework, future work might center on optimizing hyperparameters, investigating extra machine learning calculations, and refining information preprocessing strategies.

V. CONCLUSION

The study concludes that leveraging advanced machine learning algorithms significantly enhances the accuracy of insurance claim predictions. The research highlighted the effectiveness of Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) in improving prediction reliability. SVM demonstrated superior performance due to its ability to manage complex data and accurately classify claim types. The optimization and tuning of model parameters played a crucial role in achieving these high accuracy levels.

Additionally, the findings emphasize that integrating these machine learning techniques can substantially benefit the insurance industry by improving risk management and operational efficiency. Accurate predictions enable better financial planning and resource allocation, which enhances claim processing and reduces operational costs. The study underscores the practical advantages of adopting such models, suggesting that they offer valuable insights for managing insurance risks and improving customer satisfaction.

The research affirms that advanced algorithms like SVM and KNN are instrumental in refining insurance claim predictions. These methods not only improve prediction accuracy but also provide a strategic edge for insurers. The results pave the way for future developments in predictive modeling, encouraging ongoing innovation in the insurance sector.

REFERENCES

[1]. Giovanni Salzillo and Giovanni B. Fioccola “Evaluation of Driver Drowsiness based on Real- Time Face Analysis”, IEEE International Conference on Systems, Man, and Cybernetics (SMC) October 11-14, 2020. Toronto, Canada.

[2]. Igor Lashkov, Alexey Kashevnik, Nikolay Shilov, Vladimir Parfenov and Anton Shabaev “Driver Dangerous State Detection Based on OpenCV & Dlib Libraries Using Mobile Video Processing”, IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2019.

[3]. Kyong Hee Lee, Whui Kim, Hyun Kyun Choi and Byung Tae Jang “A Study on Feature Extraction Methods Used to Estimate a Driver’s Level of Drowsiness”, International Conference on Advanced Communications Technology(ICACT), February 17-20, 2019.

[4]. LAMIA ALAM, MOHAMMED MOSHIUL HOQUE, M. ALI AKBER DEWAN, NAZMUL SIDDIQUE, INAKI RANO AND IQBAL H. SARKER “Active Vision-Based Attention Monitoring System for Non-Distracted Driving”, VOLUME 9, 2021.

[5]. Mohsen Shirpour, Steven S. Beauchemin and Michael A. Bauer “A Probabilistic Model for Visual Driver Gaze Approximation from Head Pose Estimation”, IEEE 3rd Connected and Automated Vehicles Symposium (CAVS), 2020.

[6]. Panteleimon-Evangelos Aivaliotis, Foteini Grivokostopoulou, Isidoros Perikos, Ioannis Daramouskas and Ioannis Hatziligeroudis “Eye Gaze Analysis of Students in Educational Systems”, IEEE, 2020.

[7]. Vidhu Valsan A, Paul P Mathai and Ierin Babu “Monitoring Driver’s Drowsiness Status At Night Based On Computer Vision”, International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021.

[8]. Wissarut Kongcharoen, Siranee Nuchitprasitchai, Yuenyong Nilsiam and Joshua M. Pearce “Real-Time Eye State Detection System for Driver Drowsiness Using Convolutional Neural Network”, 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2020.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

[9]. Yang Xing, Huaji Wang and Efstathios Veleniso "Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach", IEEE, 2018.
[10]. Yiheng Zhao, Abdelhamid Mammeri and Azzedine Boukerche "A Novel Real-time Driver Monitoring System Based on Deep Convolutional Neural Network", IEEE 2019.