Research Article

AI Explainability in Insurance Applications with Telematics Data

Dr. Rajkumar J. Bhojan[1], Dr. Ramyachitra[2], Dr. Subra Ganesan[3], Ragavi R[4], Roshini Rajkumar[5]

[1] Principal Engineer, Fidelity Investments, Smithfield, RI, US. Email:jbraviva@gmail.com

[2] Associate Professor, Dept. of Computer Science & Engineering, Bharathiar University, Coimbatore, India.

[3] Professor, Electrical & Computer Engineering, Oakland University, Rochester, MI, USA. E-mail: ganesan@oakland.edu.

[4] Software Engineer, Fidelity Investments, RI, USA

[5] Senior Student, UMass Amherst, MA, USA

Abstract: This paper describes a step-by-step process of creating an insurance-based application that predicts insurance premium using Artificial Intelligence (AI) method of machine learning (ML). Artificial Intelligence technologies will have a high impact on all aspects of the insurance industry, from distribution to underwriting and pricing to claims. Nowadays, AI is being used in insurance claims processing, underwriting, fraud detection, anomaly detection and customer service. In this research paper, we are going to predict a driver's premium charges based on the driver's driving behavior. Here, we used telematics, machine learning, and deep learning technologies for predicting the insurance premium cost in the auto insurance space. In this research journey, we propose a hybrid framework architecture with multiple combinations of Supervised Learning (LR, DS, RFR and SVM) Algorithms and interpretability of AI model.

Keywords: Machine Learning, Supervised Learning, Insurance, Explainability, Customer Service

---

1. Introduction

Authors in [1] describe that AI is becoming the universal engine of execution. As digital technology increasingly shapes "all of what we do" and enables a rapidly growing number of tasks and processes, AI is becoming the new operational foundation of business—the core of a company's operating model, defining how the company drives the execution of tasks. AI is not only displacing human activity; it is changing the very concept of the firm. To get full potential value from AI, it requires carriers to integrate skills, technology, and insights from around the organization to deliver unique, holistic customer experiences. .

The current machine learning techniques rely on the availability of large, labeled datasets to achieve state-of-the-art results on tasks such as image classification[2], speech recognition [3], and machine translation[4] (Wu et al., 2016).

The author, Chip Huyen[5] describe that machine learning is an approach to learn complex patterns from existing data and use these patterns to make predictions on unseen data. In this research work, we used Supervised Learning algorithms for our prediction. Supervised Learning is one of the major categories in machine learning. In supervised learning, the training set you feed to the algorithm includes desired solution,

called labels.[6].

The authors in this paper explain that the process of training an ML model involves providing an ML algorithm with training data to learn from. The most used models are: binary classification model, multi class classification model, and regression model. For each of these models there are well-known learning algorithms such as decision trees, kNN (k-Nearest Neighbors, random forests, neural networks, logistic regression, linear regression, genetic algorithms, SVM (Support Vector Machine), naive Bayes, and more.[7]

2. Telematics:

According to Wikipedia, Telematics is an interdisciplinary field that encompasses telecommunications, vehicular technologies, electrical engineering, and computer science [8]. In this research, we are dealing with vehicular technologies, the insurance domain, and AI.

Authors in this paper implemented a prototype of how an end-to-end PayHowYouDrive (PHYD) solution, offering the power of Apache Spark analytics while protecting the privacy rights of the data subjects, can be securely implemented on a public cloud [9].

Authors in this paper [10] describe that in the insurance industry, the assessor's role is essential and requires significant efforts conversing with the claimant. This is a highly professional process that involves many complex operations to make a final insurance report. The authors also tried to build an NLP (Natural Language Processing) based system to help junior assessor.

The author Tony Boobier appreciates the telematics technology in his book "Analytics for Insurance". He mentioned that since 2010, the use of telematics Author Tony Boobier narrates that since 2010, the use of telematics in auto insurance has grown and it is no longer seen as a niche underwriting product for hazardous drivers. He further narrates that this technology has evolved beyond 'Pay as you drive' to "Pay how you drive" [11].

Along with this, other typical indicates might include:

- How you drive
- When you drive
- What you drive
- What color is your car
- Where you drive to
- What else you drive.

He further extends that Telematics is an example of the use of integrated technology at its best.

The below table shows some of the mobile apps being used by top Insurance Companies in the USA.

S.No.	Insurance Company	Apps Name	Tracking Factors	References
1	Geico	DriveEasy	Speeding, Braking, Phone Use, Time of day, Distance	[12]
2	USAA	SafePilot	Phone Handling and hands-free phone use, Braking	[13]
3	AllState	Drivewise	Safe speed, Braking, Time of Day, Mileage	[14]
4	StateFarm	Drive Safe & Save	Quick acceleration, Hard braking, Fast Cornering, Speeding, Annual Mileage	[15]
5	American Family Insurance	KnowYourDrive	Vehicle usage, Braking, Acceleration, Speed, Time of day, Distracted driving habits	[16]

Table-1 Insurance Apps

3. Proposed Solution

The proposed solution will collect the driver's behavior data in real-time, processed the data cleaning, train the data for the predictions.

Knowingly or unknowingly, we get an emotional distraction when we drive, maybe because of talking with peer members, songs we listen to, other driver's rash behavior on roads, high traffic jams, etc. Authors in [17] describe that driver distraction is estimated to be one of the leading causes of motor vehicle accidents. However, little is known about the role of emotional distraction on driving, despite evidence that attention is highly biased toward emotion. In their study they used a dual-task paradigm to examine the potential for driver distraction from emotional information presented on roadside billboards.

3.a Negative Emotions

The emotions may be positive or negative, As stated by DMV.org [18]
below are some examples of negative events that may affect your mood and ability to focus on driving include:

- Getting into an argument with your significant other.
- A stressful day at work.
- Another driver on the road cuts you off.
- Receiving sad/distressing news.
- Running late to an important appointment.

3.b Positive Emotions

Similar to the effects of negative emotions, positive life events can also leave you just as distracted on the road. A few examples of positive situations that could result in distracted driving include:

- Receiving a raise at work.
- Getting good/exciting news.
- Heading to or from a celebration.
- Winning a prize.
- Listening/singing along to a song you like.

Although these situations may not seem problematic at all, getting wrapped up in the emotions that come as a result could lead to very negative consequences**.** In fact, for some it might even be harder to let go of positive emotions than negative ones.

The consequences of these emotional distractions may cause accidental damages to drivers or peer drivers on the road. Some examples may include:

- Drifting into another lane.
- Driving through a red light or stop sign.
- Rear-ending the car ahead of you.
- Hitting a pedestrian.

Authors in [19] mentioned that in the automotive industry, researcher, AI experts and developers are pushing deep learning based approaches for in-vehicle safety. In their research paper, they propose a hybrid deep learning based visual system for providing feedback to the driver in a non-intrusive manner.

4. Results and Discussion:

For this research work, we implemented a proof of concept with available telematics dataset. We used “fer2013” training dataset, Angular for Front-end, FastAPI for middleware, PostgreSQL for backend and Pytest for testing the application.

The emotional distractions will be analyzed by our supervised learning methodologies. We mount a camera in front of driver's seat, which observes his facial emotions continuously on the road and sends data to a customized mobile apps mentioned in Table-1. The mobile apps collect the driver’s data continuously and stores them for further analysis. We divide the data in different integer values as shown below in Table-2.

Emotional Distractions	Values assigned
Normal – there is no distraction	0
Positive Emotions	1
Negative Emotions	2
Texting	3

Table-2 Emotional Distractions

Authors Harpreet Singh, et. al. attempted to review the naturalistic driving studies in their paper [12]. The paper begins with an overview of different methods for collecting unobtrusive driver behavior data during their day-to-day trip, followed by a discussion of various factors affecting driving behavior and their influence on vehicle performance parameters. The paper also discusses the strategies mentioned in the literature for improving driving behavior using naturalistic driving studies to enhance road safety. The paper also discusses the strategies mentioned in the literature for improving driving behavior using naturalistic driving studies to enhance road safety. Some of the major findings of this review suggest that i) driver behavior is a major cause in the majority of the road accidents ii) drivers generally reduce their speed and increases headway as a compensatory measure to reduce the workload imposed during distracting activity and adverse weather conditions iii) mobile phone has emerged as a potential device for collecting naturalistic driving data and, iv) improvement in driving behavior can be achieved by providing feedback to the drivers about their driving behavior.

Texting while driving has been highlighted as a dangerous behavior that leads to impaired judgement and accidents.

Dataset Overview:

The below graphs explain the structure of our sample dataset.

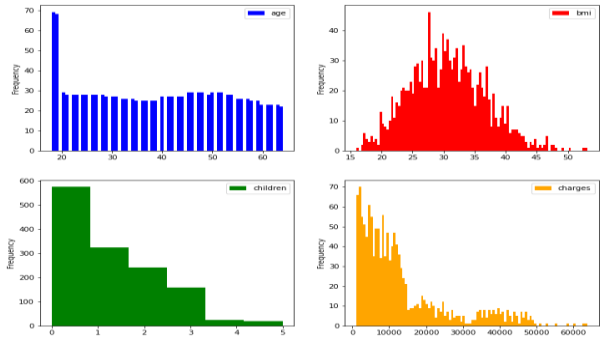


Fig-1. Age, Bmi, Children, Charges

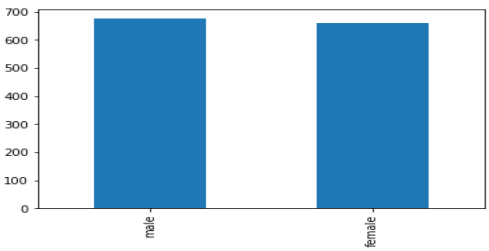


Fig-2. Male and Female Count

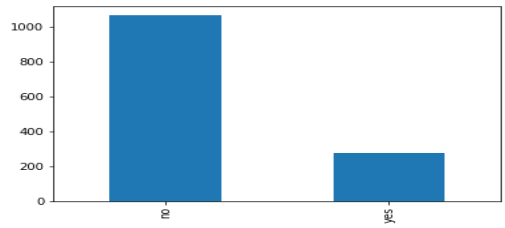


Fig-3 Smoker's Count

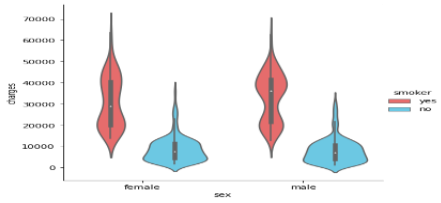


Fig-4 Smoker Vs Charges

4.a AI Interpretability in Model Evaluation:

We further extended our research with Explainability of the model and its performance. As per [13], Models are interpretable when humans can readily understand the reasoning behind predictions and decisions made by the model. It further explains that the more interpretable the models are, the easier it is for someone to comprehend and trust the model. As mentioned in the research paper [14], the authors cite that there is no established consensus on the complete set of factors that makes an AI system explainable, XAI work commonly shares the goal of making an AI system's functioning or decisions easy to understand by people. We used LIME (Local Interpretable Model-agnostic Explanations), it is a visualization technique that helps explain individual predictions. As it is a model agnostic, we used it for our supervised regression model.

As shown in the visualization, the colors blue and orange, depicting negative and positive associations of the model.

Hence, we concluded that, the following features played major contributions in predicting the insurance charges.

1. Smoker
2. age
3. bmi
4. drivedata

The other features like sex and children data did not involve much in this prediction.

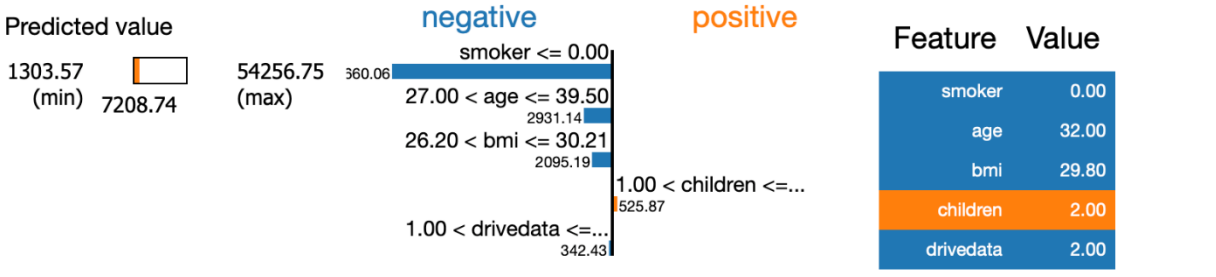


Fig-5 Model Explanations

Conclusion:

The proposed approach is based on the combination of different Supervised Learning (LR, DS, RFR, SVM) Algorithms. Our Experimental results demonstrate that hybrid approach gives the overall detection accuracy of 91.98%, better than static and dynamic detection accuracies of 71.46% and 81.13% respectively. We also learnt that if we use more data in our training, the model prediction will be higher and accurate. Moreover, our Interpretability technique gives more explainable about the model prediction. In our future research, we would be utilizing different types of Explainable AI techniques like, SHAP, GRAD-CAM, DeepLIFT, etc., for better enhancements of the model accuracy and prediction.

Acknowledgement:

The authors thank project stakeholders for their valuable support in collecting the data. I am immensely thankful to Dr. D. Ramyachitra, who provided insight and expertise that greatly assisted this research paper. I sincerely thank Dr.Subramaniam Ganesan for his constant encouragement, guidance, and support throughout this research. I also thank Ragavi. R for her help.

References:

[1] Competing in the Age of AI - Marco Iansiti and Karim R. Lakhani
[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database", Scribd, 2009.
[3] Dario Amodei, et.al "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin", arXiv:1512.02595[cs.CL], 2015.
[4] Yonghui Wu et.al "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation"
arXiv:1609.08144v2[cs.CL], 2016,
[5] Chip Huyen, "Designing Machine Learning Systems", O'reilly publications, May 2022
[6] Aurelien Geron, "Hands-on Machine Learning with Scikit-Learn, Keras, and Tensoflow", Third Edition, October 2022, O'reilly media Inc.
[7] Marcose Pinto, et.al., "Health care AI: Predicting Breast Cancer with Machine Learning" Consortium for Computer Science in Colleges, 2018.
[8] Wikipedia
[9] Eliot Salant, Gidon Gershinsky "End-to-end secure insurance telematics", Proceedings of the 12th ACM International Conference on System and Storage, May 2019.

[10] Shuange Pend, Minghui Yang, et. al, "Insurance Assistant: An Intelligent Platform for Video Insurance Assessment", DOI: 10.1145/3442442.3458600
[11] Analytics for Insurance, Tony Boobier
[12] https://www.geico.com/driveeasy
[13] https://www.usaa.com/insurance/safedriving?akredirect=true
[14] https://www.allstate.com/drivewise
[15] https://www.statefarm.com/insurance/auto/discounts/drive-safe-save
[16] https://www.amfam.com/insurance/car/knowyourdrive
[17] Michelle Chan, "The emotional side of cognitive distraction: Implications for road safety", National Library of Medicine, DOI: 10.1016/j.aap.2012.04.004
[18] DMV.org
[19] Dr. Rajkumar J Bhojan, Dr. Ramya Chitra, Dr. Subra Ganesan, Ragavi. R "A Hybrid Deep Learning Based Visual System for In-Vehicle Safety", European Journal of Engineering Research and Science, Vol-4, April 2019.
[20] Harpreet Singh, Ankit Kathuria, "Analyzing driver behavior under naturalistic driving conditions: A review,Accident Analysis & Prevention", Volume 150, 2021, 105908, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2020.105908.
[21] https://www.interpretable.ai/interpretability/what/
[22] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, Justin D. Weisz, "Expanding Explainability: Towards Social Transparency in AI systems", DOI:10.1145/3411764.3445188, ACM, May 2021.