

Erklärbarkeitsmethoden für Machine-Learning-Modelle am Beispiel des Wiederanlageverhaltens in einem Versicherungsbestand

Zusammenfassung der Bachelorarbeit an der Universität Ulm

Johanna Elise Hell

Zielsetzung und Relevanz der Arbeit

Diese Abschlussarbeit untersucht den Einsatz erklärbarer künstlicher Intelligenz (Explainable AI, XAI) im Versicherungsbereich am Beispiel der Prognose von Wiederanlageentscheidungen nach Ablauf von Lebens- und Rentenversicherungen. Ziel der Arbeit ist die Entwicklung eines maschinellen Lernmodells zur Vorhersage von Wiederanlageentscheidungen bei Ablaufleistungen sowie dessen Nutzung zur Unterstützung vertrieblicher Entscheidungsprozesse, insbesondere um gezielt zu identifizieren, welche Kundinnen und Kunden mit hoher Wahrscheinlichkeit für eine Wiederanlage angesprochen werden sollten. Ergänzend dazu wird die Nachvollziehbarkeit der Entscheidungslogik des Modells mithilfe verschiedener XAI-Methoden sichergestellt.

Die Arbeit verbindet damit aktuelle Entwicklungen aus den Bereichen Machine Learning, Explainable AI und Versicherungswirtschaft und zeigt praxisnah, wie moderne KI-Verfahren in einem hochregulierten Umfeld eingesetzt werden können. Gleichzeitig wird verdeutlicht, dass erklärbare KI nicht ausschließlich aus technischer Sicht relevant ist, sondern auch eine wichtige Rolle im Spannungsfeld zwischen Wirtschaftlichkeit, Ethik und regulatorischen Anforderungen spielt.

Die Relevanz dieses Themas ergibt sich aus dem zunehmenden Einsatz von künstlicher Intelligenz in der Versicherungswirtschaft sowie den erhöhten regulatorischen Vorgaben in dem Bereich, insbesondere durch den EU AI Act. Diese machen in einem kundenkritischen Umfeld transparente, nachvollziehbare und faire Entscheidungen unverzichtbar. XAI stellt dabei eine zentrale Voraussetzung dar, um Modellentscheidungen verständlich zu machen und ihre Anwendung im operativen Geschäft verantwortungsvoll zu gestalten.

Modellentwicklung und Evaluation

Die empirische Grundlage des Modells bildet ein realistischer Datensatz eines deutschen Lebensversicherers mit über 134.000 Verträgen und 27 aufbereiteten Merkmalen zum abgelaufenen Vertrag und zum Kunden. Die binäre Zielvariable beschreibt, ob innerhalb von 15 Monaten nach Vertragsabschluss eine Wiederanlage der Ablaufleistung erfolgt. Der Datensatz wird umfassend bereinigt, reduziert und hinsichtlich Klassenungleichgewichten sowie Ausreißern aufbereitet.

Im Rahmen der Datenvorverarbeitung werden unter anderem Duplikate entfernt, hoch korrelierte Variablen reduziert sowie seltene Merkmalsausprägungen ausgeschlossen, um die Stabilität und Robustheit des Modells zu verbessern. Zusätzlich werden Extremwerte mithilfe eines Winsorizing-Verfahrens begrenzt, sodass Ausreißer die Modellierung weniger stark beeinflussen.

Zur Modellierung wird ein Gradient-Boosting-Verfahren eingesetzt, das durch die Kombination vieler schwacher Entscheidungsbäume komplexe nichtlineare Zusammenhänge abbildet. Die Modellleistung wird anhand gängiger Klassifikationsmetriken wie ROC-AUC, Precision, Recall und F1-Score bewertet und unter Berücksichtigung der starken Klassenungleichverteilung im Datensatz eingeordnet.

Die Ergebnisse zeigen, dass das Modell eine gute Trennschärfe zwischen den Klassen erreicht und Wiederanlageentscheidungen deutlich besser als ein zufälliger Klassifikator vorhersagen kann. Dennoch treten weiterhin Fehlklassifikationen auf, insbesondere in Form von False Positives und False Negatives, was auf weiteres Optimierungspotenzial hinsichtlich der Modellgüte hinweist.

Die Analyse zeigt außerdem, dass die Wahl geeigneter Schwellenwerte einen erheblichen Einfluss auf die Bewertung des Modells besitzt. Abhängig von der vertrieblichen Zielsetzung kann entweder eine höhere Trefferquote potenzieller Wiederanlagen oder eine Reduktion fehlerhafter Vorhersagen priorisiert werden. Dadurch wird deutlich, dass die Bewertung von KI-Modellen stets im konkreten Anwendungskontext erfolgen muss.

Globale und lokale Erklärbarkeit der Modellentscheidungen

Zur Interpretierbarkeit des Modells werden sowohl globale als auch lokale XAI-Methoden eingesetzt, um die Entscheidungslogik sowie die zugrunde liegenden Einflussfaktoren der Vorhersage einer Wiederanlage transparent zu machen. Ziel ist es insbesondere, die Merkmale zu identifizieren, die den Modelloutput maßgeblich bestimmen und die Prognose entscheidend beeinflussen. Globale Verfahren wie Feature Importance, Partial Dependence Plots und ICE identifizieren zentrale Einflussfaktoren auf die Modellvorhersage, insbesondere den Vorsorgetyp, die Vertragslaufzeit und die Ablaufleistung. Lokale Methoden wie LIME und SHAP ermöglichen die detaillierte Erklärung einzelner Vorhersagen und zeigen, dass Entscheidungsprozesse stark von individuellen Merkmalskombinationen und Interaktionen geprägt sind. Insgesamt liefern die eingesetzten XAI-Methoden umfassende Einblicke in die Entscheidungsstruktur des Modells, deren Interpretation jedoch stets unter Berücksichtigung der methodischen Annahmen und Grenzen erfolgen muss.

Die lokalen Erklärungen bestätigen zentrale Ergebnisse der globalen Analyse. So wirken insbesondere eine lange Vertragslaufzeit sowie ein höheres Alter der versicherten Person zum Vertragsende tendenziell negativ auf die Wiederanlagewahrscheinlichkeit, während eine hohe Ablaufleistung und bestimmte Haushaltsstrukturen diese positiv beeinflussen. Diese Zusammenhänge zeigen sich konsistent sowohl auf globaler als auch auf individueller Ebene.

Gleichzeitig treten Unterschiede in der Bedeutung einzelner Merkmale je nach Betrachtungsebene auf, da lokale Erklärungen zusätzliche einzelfall-spezifische Einflussfaktoren sichtbar machen, die in der aggregierten Analyse weniger stark hervortreten. Insgesamt ergänzen sich beide Perspektiven und ermöglichen ein umfassendes und nachvollziehbares Verständnis der Modellentscheidungen.

Darüber hinaus zeigt die Arbeit, dass globale und lokale Erklärungsverfahren unterschiedliche Stärken besitzen. Während globale Methoden insbesondere allgemeine Muster und Zusammenhänge im Datensatz sichtbar machen, ermöglichen lokale Verfahren die detaillierte Analyse individueller Modellentscheidungen. Erst die Kombination beider Perspektiven erlaubt eine ganzheitliche Interpretation des Modellverhaltens.

Gleichzeitig werden auch die Grenzen erklärbarer KI deutlich. Insbesondere Methoden wie LIME und SHAP basieren teilweise auf vereinfachenden Annahmen oder künstlich erzeugten Datenpunkten, wodurch Erklärungen unter bestimmten Bedingungen instabil oder schwer interpretierbar sein können. Die Ergebnisse der XAI-Methoden müssen daher stets kritisch eingeordnet und fachlich bewertet werden.

Fazit

Insgesamt zeigt die Arbeit, dass erklärbare künstliche Intelligenz einen zentralen Beitrag zur transparenten und nachvollziehbaren Modellierung von Wiederanlageentscheidungen im Versicherungswesen sowie in weiteren Anwendungsbereichen der Versicherungswirtschaft leisten kann. Die Kombination eines leistungsfähigen Gradient-Boosting-Modells mit globalen und lokalen XAI-Verfahren ermöglicht es, sowohl allgemeine Einflussstrukturen als auch individuelle Entscheidungslogiken systematisch zu analysieren.

Gleichzeitig bestätigt die Analyse, dass globale und lokale Interpretationsansätze komplementäre Perspektiven auf das Modellverhalten liefern und erst in ihrer Kombination ein umfassendes Verständnis der Entscheidungsstruktur ermöglichen. Die Ergebnisse unterstreichen zudem die Notwendigkeit, maschinelle Lernverfahren im Versicherungskontext nicht nur hinsichtlich ihrer Prognosegüte, sondern insbesondere auch hinsichtlich ihrer Nachvollziehbarkeit und Transparenz zu bewerten.

Darüber hinaus verdeutlicht die Arbeit die Relevanz erklärbarer KI im Spannungsfeld zwischen analytischer Leistungsfähigkeit, regulatorischen Anforderungen und ethischen Aspekten automatisierter Entscheidungsprozesse. Insbesondere im Versicherungswesen ist Transparenz von entscheidender Bedeutung, um sicherzustellen, dass Entscheidungen nicht ausschließlich auf Basis intransparenter Modelle getroffen werden, sondern nachvollziehbar und verantwortungsvoll gestaltet sind. Auf diese Weise leisten XAI-Methoden einen wesentlichen Beitrag zur Reduktion von Black-Box-Charakteristika und schaffen die Grundlage für vertrauenswürdige und praxisnahe Anwendungen im Versicherungsbereich.

Die entwickelten Ansätze weisen darüber hinaus ein erhebliches Potenzial für vielfältige weitere Anwendungsfelder innerhalb der Versicherungswirtschaft auf. Auch außerhalb des Versicherungssektors gewinnen Methoden der erklärbaren Künstlichen Intelligenz zunehmend an Bedeutung, insbesondere in Bereichen, in denen sensible oder regulatorisch relevante Entscheidungsprozesse durch automatisierte Systeme unterstützt werden.

Zusammenfassend verdeutlicht die Arbeit, dass erklärbare künstliche Intelligenz eine wesentliche Grundlage für vertrauenswürdige, regulatorisch konforme und praxisnahe KI-Anwendungen darstellt. Durch die Verbindung leistungsfähiger Machine-Learning-Modelle mit transparenten Erklärungsverfahren kann ein wichtiger Beitrag zur verantwortungsvollen Nutzung künstlicher Intelligenz in der Versicherungswirtschaft geleistet werden.