

Customer Churn Prediction - Ein CAT-Bericht

Volkan Korunan

8. November 2024

Konzept

Einführung in Data Science - Customer Churn Prediction

Thema: Anfertigen eines CAT für die Vorhersage der Kundenabwanderung

Einleitung

Kundenabwanderung stellt für Unternehmen eine ernsthafte Herausforderung dar, da der Verlust von Kunden langfristig zu Umsatzrückgängen und erhöhten Akquisitionskosten führt. Durch die Identifizierung abwanderungsgefährdeter Kunden können Unternehmen gezielte Maßnahmen ergreifen, um diese Kunden zu halten und die Kundenzufriedenheit zu steigern.

Ziele dieser Arbeit

Das Ziel dieser Arbeit ist es, ein Modell zur Vorhersage der Kundenabwanderung zu entwickeln, das eine höhere Genauigkeit als das bestehende ANN-Modell auf Kaggle erreicht, welches eine Genauigkeit von 87% erzielt hat. Der Ansatz dieser Arbeit verwendet ein Ensemble-Modell, bestehend aus XGBoost, LightGBM und CatBoost, um die Vorhersageleistung zu verbessern.

Begründung des Themas

Warum das Thema wichtig ist

- Kundenabwanderung stellt für Banken ein großes wirtschaftliches Risiko dar, da der Verlust von Bestandskunden erhebliche Umsatzeinbußen be-

deutet. - Frühzeitige Erkennung abwanderungsgefährdeter Kunden ermöglicht gezielte Maßnahmen zur Stärkung der Kundenbindung.

Warum maschinelle Lerntechniken dafür angewendet werden

- **Erkennung komplexer Muster:** Maschinelle Lerntechniken können komplexe Beziehungen in Kundendaten erkennen, die traditionelle Methoden nicht erfassen.
- **Hohe Vorhersagegenauigkeit:** Ensemble-Modelle können die Genauigkeit der Vorhersagen erheblich steigern, was für präzise Maßnahmen entscheidend ist.
- **Effizienz:** Automatisierte Datenverarbeitung spart Ressourcen und beschleunigt die Entscheidungsfindung.
- **Anpassungsfähigkeit:** Modelle können mit neuen Daten kontinuierlich verbessert werden.

Darstellung eines persönlichen Erkenntnisinteresses

Ich finde die Untersuchung zur Vorhersage der Kundenabwanderung durch Machine Learning faszinierend, da sie Unternehmen hilft, die Kundenbindung zu verbessern. Fortschritte in der Anwendung maschineller Lernmodelle könnten global dazu beitragen, den Erfolg von Unternehmen zu sichern.

Wofür ist das Ensemble-Modell?

Das Ensemble-Modell verbessert die Vorhersagegenauigkeit, indem es mehrere Modelle kombiniert. Es bietet eine robustere und zuverlässigere Lösung, ideal für die Vorhersage von Kundenabwanderung.

Ergebnisse

Darstellung der Ergebnisse

Modellgenauigkeit: 90%, ROC AUC Score: 0,96.

Abbildung 1 zeigt die Konfusionsmatrix des Modells. Abbildung 2 veranschaulicht die ROC-Kurve, Abbildung 3 die Feature-Importance und Abbildung 4 die Wahrscheinlichkeitsverteilung.

Interpretation der Ergebnisse

Das Modell zeigt eine gute Balance zwischen Präzision und Recall. Die hohe AUC-Wertung deutet auf eine starke Trennschärfe hin.

Ausblick und Verbesserungsvorschläge

Das Modell erreicht eine hohe Genauigkeit, jedoch bleibt das Ziel, über 98% zu erreichen. Zukünftige Arbeiten sollten sich auf die Optimierung der Hyperparameter und die Minimierung der False Negatives konzentrieren.