Guten Tag, liebe Mitarbeiterinnen und Mitarbeiter von FlexAppealFitness!

Ich freue mich sehr, Ihnen heute unsere Ergebnisse zur Vorhersage von Vertragskündigungen („Churn“) präsentieren zu dürfen. Unser Ziel war es, zu überprüfen, wie gut wir anhand Ihrer bestehenden Kundendaten vorhersagen können, welche Mitglieder ihren Vertrag wahrscheinlich beenden werden – und wie wir diese Vorhersage mit einer ganz einfachen Grundannahme, einer sogenannten „Baseline“, vergleichen.

**1. Was ist unsere Baseline und warum haben wir sie gewählt?**

Zunächst einmal haben wir uns gefragt: Wie würde ein Modell abschneiden, das gar keine Informationen nutzt, sondern einfach annimmt, dass **niemand** kündigt? Diese Herangehensweise nennen wir „Baseline“. Konkret bedeutet das: Wir sagen für **jeden** Kunden vorher, dass er seinen Vertrag nicht kündigen wird. Auf Basis Ihrer historischen Daten hatte diese Baseline einen **Fehleranteil von 26,525 %**. Das heißt: Rund 27 % der tatsächlichen Kündiger würden mit dieser einfachen Annahme nicht erkannt – und diese Fehlerquote dient uns nun als Referenzwert. Jeder Modellierungsansatz, der hier deutlich besser abschneidet, ist bereits ein Fortschritt.

**2. Unsere Machine-Learning-Modelle und ihre Leistung**

Anschließend haben wir vier verschiedene Klassifikationsmodelle trainiert und verglichen:

1. **Random Forest**
2. **Gradient Boosted Tree**
3. **k-Nearest Neighbors (kNN)**
4. **Logistische Regression**

Alle vier Modelle haben gezeigt, dass sie Ihre Kundendaten sehr viel besser nutzen können als die Baseline „niemand kündigt“. Nach eingehender Prüfung haben wir festgestellt, dass die **Logistische Regression** das beste Ergebnis geliefert hat. Lassen Sie mich kurz erklären, was das für uns bedeutet:

* Die **Fehlerquote** der Baseline lag bei **26,525 %**.
* Unsere **Logistische Regression** hat eine **Fehlerquote von etwa 7 %** erreicht.
* Das entspricht einer **Genauigkeit** (Accuracy) von **93 %**.

Damit sind wir deutlich besser als nur „niemand kündigt“: Statt etwa 27 % falscher Vorhersagen im Schnitt haben wir nur noch rund 7 %. Anders formuliert: Von 100 tatsächlichen Kündigungen erzielen wir statt 73 % lediglich knapp 7 % Fehlklassifikationen, der Rest wird korrekt als Kündigung erkannt.

**3. Vergleich der Modelle im Überblick**

Um Ihnen eine Vorstellung zu geben, wie die einzelnen Methoden abschneiden, hier ein kurzer Vergleich ohne zu viele technische Details:

* **Random Forest**: Liefert robuste und relativ stabile Ergebnisse, weil viele Entscheidungsbäume kombiniert werden. Die Fehlerquote lag im Bereich von etwa durchschnittlichen 18 %.
* **Gradient Boosted Tree**: Verbessert die Vorhersage schrittweise, indem jeder nachfolgende Baum die Fehler des vorherigen korrigiert. Das Ergebnis war ähnlich gut wie bei Random Forest, mit einer durchschnittlichen Fehlerquote um die 9 %.
* **k-Nearest Neighbors (kNN)**: Klagt manchmal, wenn die Datenpunkte sehr unterschiedlich verteilt oder Datenmengen groß sind. Wir lagen hier um die 8 % Fehler im Durchschnitt, was aber noch okay sein kann.
* **Logistische Regression**: War durchweg stabil, gut interpretierbar und erreichte im Test-Datensatz die **niedrigste durchschnittliche** **Fehlerquote mit etwa 7 %**.

Alle Verfahren liegen also deutlich unter der Baseline­-Fehlerquote von 26,5 %, doch die Logistische Regression war durchweg am zuverlässigsten.

**4. Warum die Logistische Regression für FlexAppealFitness besonders geeignet ist**

Sie fragen sich vielleicht, warum gerade die Logistische Regression für Sie den Spitzenplatz eingenommen hat. Drei Gründe sprechen besonders dafür:

1. **Einfachheit und Nachvollziehbarkeit**  
   Die Logistische Regression ist leichter zu erklären und zu verstehen. Wir können transparent zeigen, welche einzelnen Merkmale besonders stark mit einer Kündigung zusammenhängen. Zum Beispiel erkennen wir, ob eine niedrige Kursfrequenz oder hohe Zusatzkosten besonders stark auf eine Kündigung hindeuten.
2. **Gute Performance bei gemischten Daten**  
   Ihre Daten enthalten sowohl binäre Merkmale (wie „Promo durch Freunde“ oder die Angaben zum Telefon) als auch durchschnittliche Werte wie Zusatzkosten oder Kursfrequenz. Die Logistische Regression kommt mit dieser Mischung sehr gut zurecht und liefert konstante Ergebnisse.
3. **Schnelle Trainingszeiten und einfache Wartung**  
   Sollte sich Ihr Datensatz in den nächsten Monaten noch weiter vergrößern, skaliert die Logistische Regression problemlos. Sie trainiert in der Regel schneller als komplexere Ensembleverfahren wie Random Forest oder Gradient Boosted Tree. Wenn Sie also regelmäßig neue Modelle auf Ihren aktuellen Daten erstellen möchten, ist das ein klarer Vorteil.

**5. Automatisiertes Hyperparameter-Tuning**

Ein wichtiger Baustein bei allen unseren Modellen ist das **Hyperparameter-Tuning** – also die Suche nach den besten Grundeinstellungen, um die Vorhersagequalität zu maximieren. Dieser Prozess läuft in KNIME vollständig **automatisiert** ab:

* Wir legen für jedes Modell einen Bereich an Parametern fest, die wir variieren möchten.
* KNIME nimmt diese Parameterbereiche und führt für jede Kombination eine kleine „innere“ Kreuzvalidierung durch.
  + Das bedeutet, dass das Trainings-Set in mehrere Teilmengen aufgeteilt und mehrfach durchprobiert wird, um zu sehen, bei welchen Einstellungen das Modell im Durchschnitt am besten ist.
* Am Ende wird automatisch die Einstellung ausgewählt, die in der Cross-Validation am besten abgeschnitten hat – und diese Parameter geben wir an das finale Modelltraining weiter.

Dank dieses automatischen Tuning-Prozesses können wir sicher sein, dass unserer Modelle wirklich mit optimalen Grundeinstellungen versehen sind, ohne stundenlang per Hand probieren zu müssen.

**6. Was bedeutet das konkret für FlexAppealFitness?**

* Sie erhalten ein Modell, das **in 93 % der Fälle** korrekt voraussagt, ob ein Kunde kündigt oder nicht.
* Im Vergleich zur Baseline sparen Sie sehr viele Fehleinschätzungen ein:
  + Statt etwa 27 % falscher Vorhersagen haben wir nur noch rund 7 %.
  + Das heißt, statt 2.700 Fehleinschätzungen bei 10.000 Kunden bleiben nur noch etwa 700 übrig.
* Dadurch können Sie gezielte Maßnahmen ergreifen:
  + Persönliche Anrufe, E-Mail-Aktionen oder Sonderangebote richten Sie nur an wirklich gefährdete Kunden.
  + Ihre Marketing- und Serviceteams arbeiten effizienter, weil sie nicht mehr unzählige Bremsklötze bearbeiten müssen, sondern genau die richtigen Mitglieder erreichen.

So verbessern Sie nicht nur die Kundenbindung, sondern sparen gleichzeitig Zeit und Budget ein.

**7. Beispielhafter Praxisnutzen**

Stellen Sie sich vor, Sie haben **10 000 aktive Mitglieder**, und historisch gesehen kündigen davon etwa **2 650** im Jahr. Mit der Baseline „niemand kündigt“ würden Sie alle 2 650 Kündigungen verpassen. Unser Modell mit 7 % Fehlerquote bedeutet, dass Sie nur etwa **700 Fehlklassifikationen** haben – also etwa 700, die Sie entweder unnötig kontaktieren oder übersehen. Im Vergleich zur Baseline sind das **rund 1 950 weniger Fehler**. Das sind fast 2 000 zusätzliche Chancen pro Jahr, aktiv auf echt gefährdete Mitglieder zuzugehen und sie zu halten.

**8. Zusammenfassung und Ausblick**

* **Baseline („niemand kündigt“) hat 26,525 % Fehler**.
* **Logistische Regression erreicht 7 % Fehlerquote durchschnittlich** und damit **93 % Accuracy**.
* **Random Forest**, **Gradient Boosted Tree** und **kNN** liegen im Fehlerbereich von durchschnittlichen **8–18 %**, sind also ebenfalls deutlich besser als die Baseline, erreichen aber nicht ganz die Spitzenleistung der Logistischen Regression.
* Das **automatisierte Hyperparameter-Tuning** in KNIME stellt sicher, dass wir immer die bestmöglichen Grundeinstellungen verwenden, ohne manuellen Aufwand.
* Dadurch sparen Sie auf 10 000 Mitgliedern fast 2 000 Fehleinschätzungen pro Jahr ein – ein deutlicher Wettbewerbsvorteil.

In den nächsten Schritten könnten wir gemeinsam mit Ihnen das Modell in Ihren laufenden Betrieb integrieren: regelmäßige Updates und Monitoring, damit auch in sechs Monaten oder einem Jahr die Vorhersagen weiterhin so gut bleiben, selbst wenn sich Ihre Mitgliederstruktur ändert.

Ich bedanke mich für Ihre Aufmerksamkeit und stehe nun gerne für Fragen zur Verfügung!