**Cold calls for auto insurance**

**Problem Statement-**

The auto insurance industry faces significant challenges when it comes to employing cold calls as a marketing strategy. One of the foremost issues is the low conversion rates associated with this approach. Many individuals find unsolicited calls intrusive, leading to a negative customer experience and diminishing the likelihood of successful conversions. Moreover, the time-consuming nature of cold calling poses a substantial challenge for auto insurance companies. Reaching a sizable audience requires considerable effort and resources, often yielding modest results.

Compliance with regulations adds another layer of complexity. The auto insurance industry is subject to stringent rules governing cold calling practices, and non-compliance can result in legal consequences and financial penalties. In addition, the limited targeting capability of cold calls contributes to a lower probability of reaching individuals genuinely interested in auto insurance. The market is saturated with competitors employing similar strategies, making it challenging for companies to differentiate themselves.

Changing consumer behavior further compounds the issue, as modern individuals increasingly prefer digital communication channels over traditional phone calls. Data privacy concerns add an additional barrier, as customers may be hesitant to share personal information over the phone. Advancements in marketing technology provide alternative, more targeted approaches, rendering traditional cold calling less effective and cost-efficient.

To address these challenges, auto insurance companies must pivot towards more customer-centric and personalized marketing strategies. Leveraging technology, adhering to regulations, and adopting data-driven approaches can enhance the effectiveness of outreach efforts, ensuring a more fruitful engagement with potential customers in the ever-evolving landscape of the auto insurance industry.

**About Dataset**

**Introduction**

Here you find a very simple, beginner-friendly data set. No sparse matrices, no fancy tools needed to understand what's going on. Just a couple of rows and columns. Super simple stuff.  
As explained below, this data set is used for a competition. As it turns out, this competition tends to reveal a common truth in data science: KISS - Keep It Simple Stupid

What is so special about this data set is, given it's simplicity, it pays off to use "simple" classifiers as well. This year's competition was won by a C5.0 . Can you do better?

**Description**

We are looking at cold call results. Turns out, same salespeople called existing insurance customers up and tried to sell car insurance. What you have are details about the called customers. Their age, job, marital status, whether the have home insurance, a car loan, etc. As I said, super simple.

What I would love to see is some of you applying some crazy XGBoost classifiers, which we can square off against some logistic regressions. It would be curious to see what comes out on top. Thank you for your time, I hope you enjoy using the data set.

**Acknowledgements**

Thanks goes to the Decision Science and Systems Chair of Technical University of Munich (TUM) for getting the data set from a real world company and making it available to be shared publicly. Also Vladimir Fux, who oversees the challenge associated with this data set.

**Inspiration**

This is a data set used for teaching entry level data mining skills at the TUM. Every year there is a competition as part of the curriculum of a particular course. This Data Mining Cup teaches some of the very fundamentals that are always worthy to be revisited, especially by pros abundant at Kaggle. For some of my thoughts see the verbose comments in the Kernel.

**Dataset Link**

[Car Insurance Cold Calls (kaggle.com)](https://www.kaggle.com/datasets/kondla/carinsurance?resource=download)

**Data Correlation -**

The features exhibit a generally independent relationship, with the exception of DaysPassed and PreAttempts. The success of cold calls shows a positive correlation with PreAttempts, DaysPassed, Age, and Balance, while displaying a negative correlation with default, HHInsurance, CarLoan, LastContactDay, and NoOfContacts.

A screenshot of a chart

Description automatically generated

**Data Exploration and Findings -**

* **Age**: It's interesting to see that seniors are more likely to buy car insurance.
* **Balance**: For balance, the data point at the upper right corner might be an outlier
* **HHInsurance**: Households insured are less likely to buy car insurance.
* **CarLoan**: People with car loans are less likely to buy.
* **NoOfContacts**: Too many contacts cause customer attrition.
* **DaysPassed**: It looks like the more day passed since the last contact, the better
* **PrevAttempts**: Also, more previous attempts, less likely to buy.

According to the trained dataset, it seems that individuals aged 30 years or younger and seniors show a higher inclination to purchase car insurance from this bank.

A graph of a graph

Description automatically generated with medium confidence

**Additional Findings –**

* Job: Students are most likely to buy insurance, followed by retired and unemployed folks. This is aligned with the age distribution. There might be some promotions targeting students.
* Marital status: Married people are least likely to buy car insurance. Opportunities for developing family insurance business.
* Education: People with higher education are more likely to buy
* Communication: No big difference between cellular and telephone
* Outcome in previous campaign: Success in previous marketing campaign is largely associated with success in this campaign.
* Contact Month: Mar, Sep, Oct, and Dec are the hot months. It might be associated with school season.

**Model Evaluated On Data Set**

* K-Nearest Neighbors (KNN
* Logistic Regression
* Naive Bayes Classifier
* Random Forest

**Model Accuracy –**

* K-Nearest Neighbors (KNN) – 0.76
* Logistic Regression – 0.79
* Naive Bayes Classifier – 0.71
* Random Forest – 0.81

**Analysis**

The process of scrutinizing and constructing models reveals insightful patterns within the dataset. The predictive models, trained on historical data, establish a strong correlation between specific derived attribute features and the target variable. These models demonstrate a notable accuracy, reaching approximately 80%, underscoring their efficacy in making precise predictions. Additionally, an examination of feature importance sheds light on crucial factors influencing the model's decisions, furnishing actionable insights for decision-makers. To sum up, the analysis suggests that the developed model stands as a reliable tool for predicting the effectiveness of a cold call marketing strategy. A central concern revolves around the observed low conversion rates associated with this approach, contributing to informed decision-making and enriching our comprehension of the inherent patterns within the data.

Various Machine Learning models were experimented with for accuracy, and upon comparing the results, K-Nearest Neighbors (KNN) – 0.76, Logistic Regression – 0.79, Naive Bayes Classifier – 0.71, it becomes evident that Random Forest reports superior accuracy and is the preferred choice. Its ability to handle diverse data types, address missing values, and deliver robust performance with minimal hyperparameter tuning renders it a favored option across a spectrum of machine learning applications.

**Derived Attributes**

* Client attributes: Age, Job, Marital, Education, Default, Balance, HHInsurance, CarLoan
* Communication attributes: LastContactDay, LastContactMonth, CallStart, CallEnd, Communication, NoOfContacts, DaysPassed.
* Previous campaign attributes: PrevAttempts, Outcome

**Consider the following suggestions for enhancing the model:**

* Engage in additional feature engineering, which includes exploring interaction and polynomial terms.
* Visualize decision boundaries for select classifiers.
* Incorporate additional base models for learning.
* Experiment with different assembling approaches.