**Project Proposal (Section 3 - Team Supreme)**

**Predicting the users interested in Insurance Policy**

**Business Problem:**

Direct mailings to a company's potential customers - "junk mail" to many - can be an extremely effective way of marketing a product or service, especially in the insurance field. However, as we all know, the vast majority of people who receive junk mail are uninterested in it. If a company sends out business mail too frequently, many customers may find it irritating. Sustainability is becoming an increasingly important part of not only people's lives but also industry standards. Sending mail to thousands, if not lakhs, of people is not exactly an eco-friendly way of life, as some of the recipients may find it wasteful and harmful to the environment.

Direct mail has fixed costs because it must be routed through the system. You must also design the mail piece, which may incur additional expenses. To determine whether direct mail is a viable option for you, calculate the ROI (return on investment).

Graphical user interface, application

Description automatically generated

Most of it is thrown away, wasting the company's money, filling landfill waste sites, or needing to be recycled.

In order to avoid some of this waste and expenditure, the organization would know more precisely whom to send it to if it had a decent understanding of its potential consumers.

Therefore, if we can solve the below-mentioned questions by using a predictive model, then we would be able to reduce to the capital burden on the marketing of an insurance company and increase the efficiency of the marketing strategies.

1. To identify the potential customers who would be interested in a caravan insurance policy.

2. Describe your current or potential customers and, if possible, explain why they purchased a caravan policy.

Shape

Description automatically generated

Here, we want to develop a machine-learning model capable of identifying which customers of a Dutch insurance company have an insurance policy covering a caravan. (Caravans are mobile homes that are normally towed by cars. They are called trailers in North America.)

**Source:**

The data was supplied by the Dutch data mining company Sentient Machine Research. Information about the insurance company customers consists of 86 features and includes product usage data and socio-demographic data derived from zip area codes.

This data set can be found by clicking on this link [here](https://archive.ics.uci.edu/ml/datasets/Insurance+Company+Benchmark+(COIL+2000)).

**About Dataset:**

The dataset contains information about customer’s product usage data and socio-demographic data derived from zip area codes.

The training set contained information on 5822 customers, of which 348, about 6%, had caravan policies. The test set contained information on 4000 customers randomly drawn from the same population. For each customer, the values of 85 features were given.

We need to identify how many of the test customers were most likely to have a caravan policy.

|  |  |
| --- | --- |
| Data Description | Data Description |
| PWAPART | Contribution private third-party insurance |
| PWABEDR | Contribution third party insurance -agriculture |
| PPERSAUT | Contribution car policies |
| PBESAUT | Contribution delivery van policies |
| PMOTSCO | Contribution motorcycle/scooter policies |
| PVRAAUT | Contribution lorry policies |
| PAANHANG | Contribution trailer policies |
| PTRACTOR | Contribution tractor policies |
| PWERKT | Contribution |
| PBROM | Contribution moped policies |
| PLEVEN | Contribution life insurances |
| PPERSONG | Contribution private accident insurance policies |
| PGEZONG | Contribution family accidents insurance policies |
| PWAOREG | Contribution disability insurance policies |
| PBRAND | Contribution fire policies |
| PZEILPL | Contribution surfboard policies |
| PPLEZIER | Contribution boat policies |
| PFIETS | Contribution bicycle policies |
| PINBOED | Contribution property insurance policies |
| PBYSTAND | Contribution social security insurance polices |
| AWAPART | Number of private third party insurance |
| AWABEDR | Number of third party insurance (firms) |
| AWALAND | Number of third party insurance (agriculture) |
| APERSAUT | Number of car policies |
| ABESAUT | Number of delivery van policies |
| AMOTSCO | Number of motorcycle/scooter policies |
| AVRAAUT | Number of lorry policies |
| AAANHANG | Number of trailer polices |
| ATRACTOR | Number of tractor policies |
| AWERKT | Number of agricultural machines policies |
| ABROM | Number of moped policies |
| ALEVEN | Number of life insurances polices |
| APERSONG | Number of private accident insurance policies |
| AGEZONG | Number of family accidents insurance policies |
| AWAOREG | Number of disability insurance policies |
| ABRAND | Number of fire policies |
| AZEILPL | Number of surfboard policies |
| APLEZIER | Number of boat policies |
| AFIETS | Number of bicycle policies |
| AINBOED | Number of property insurance policies |
| ABYSTAND | Number of social security insurance policies |
| CARAVAN | Number of mobile home policies |
| Socio demographic details of customers | 43 variables related to customer zip code, area etc.,. |

**Importance of solving this problem:**

Email is one of the most important mediums for converting users into clients. Email marketing is one of the most measured and effective kinds of marketing. Spamming consumers who are not interested in insurance policies with emails will cost the firms a lot of money. On average, most businesses spend billions of dollars on emails. According to Radicati Research Group Inc., the cost of spam is not insignificant. In 2012, email spam cost organizations $20.5 billion in lost productivity and technology expenditures.

Diagram

Description automatically generated

For solving this problem, we came with analysis to predict which users are likely to be interested in a caravan insurance policy. So, that companies can send emails to interested potential customers and also be cost-effective. If the company uses this model indicates a substantial return on investment.

**Motivation:**

**Effective marketing:**

* With effective marketing, the insurance organization can increase its productivity by targeting only productive customers who are more probable to take the insurance policy.
* The insurance companies can redesign their policies who are isolated from their network by considering their interests.

**Pricing Strategy and Solution design:**

* Insurance companies must need a business model to get to know if the people are not interested in their policies due to overpricing or not.

**Real-world Impact of solving this problem:**

**Policy Optimization:**

* With predictive analytics, insurers can now create policy plans by utilizing detailed consumer data. By understanding consumer interests, value consciousness, and behavioural indications through the analysis of historical data, insurers may better match policy details and costs to what customers expect.
* Lowering the marketing costs by identifying the most successful market strategy and identifying the finest chances for future development and profit.

**Dynamic Customer Engagement:**

* The application of predictive analytics in the insurance industry is extensive when it comes to improving the client experience. Companies can utilize predictive analytics in a variety of ways to increase customer engagement. This model provides insurance firms with real-time modelling to help them a better understanding of their customers’ changing behaviours and needs.

**Methodology:**

**The figure below shows the steps that we would follow for the analysis of data –**

F1 score is more useful than accuracy, especially if the class distribution is unequal. Accuracy works best when the cost of false positives and false negatives is comparable. If the cost of false positives and false negatives differ significantly, it is preferable to include both Precision and Recall.

**Approach:**

* We will identify customers who are likely to buy a caravan policy, by developing a model which in turn, helps the insurance company know the same and thereby send mail only to those restricted customers.
* Eventually, we will recommend optimized marketing strategies to target potential customers so that they can reduce the cost incurred for marketing and manual workload.
* From the existing model features, we will engineer new features by performing feature engineering.
* After implementing the model, we will try to deploy our model and make it more efficient by making real-time predictions.

**Key Performance Indicator (KPI)**

* Key performance indicators/ Key metrics help to bridge the gap between business  
  growth and performance of our machine learning model, helping organizations function  
  efficiently towards objectives.
* The metrics which are used to evaluate our ML model are

|  |  |  |
| --- | --- | --- |
|  | **Predicted No** | **Predicted Yes** |
| **Actual No** | **TRUE NEGATIVE:** Customers who are interested in buying a caravan insurance policy, but the model predicted customers are not interested in buying the policy. | **FALSE POSITIVE:** Customers who are not interested in buying the policy and model predicted as Customers are interested in buying the policy. |
| **Actual Yes** | **FALSE NEGATIVE:** Customers who are interested to buy an insurance policy and the model predicted customers do not buy the insurance policy. | **TRUE POSITIVE:** Customers who are likely to buy an insurance policy and model predicted as customers are likely to buy an insurance policy. |

* **False Positive:** Here the customers are not interested in buying the insurance policy, but the model predicted as customers are interested in buying the policy.
* Huge investments are involved in promoting insurance company policies using email marketing. The marketing teams will target customers who are not really interested in buying the policy, ultimately leading to a waste of both time and money.
* **False Negative:** Here the customers are interested to buy insurance policies, but the model predicted customers do not buy the insurance policy.
* As a result, the insurance company suffers loss because our model is unable to predict the set of customers who are interested and will eventually lose more of their customers.
* In order to maintain customer advocacy and curb losses, we prefer false negatives over false positives.
* As False positives have less loss when compared with False Negatives, we will use Recall as our evaluation metric.
* **Note - We can consider the model with good recall value as the best fit for our data set selected.**

**Previous Analysis:**

This dataset's previous proposed solution is a data mining strategy to anticipate future clients interested in insurance policies.

* One of the contributors to the project used the Naive Bayes approach, all sociodemographic variables derived from zip codes were discarded and 2 derived variables were introduced.
* Some of the contributors used backpropagation neural networks, self-organizing maps (SOMs), evolutionary algorithms, C4.5, CART were also utilized.
* Lastly, they used the Accuracy scores derived as their main evaluation metrics.