A Life Insurance Policy Bundling Recommendation System

Masters Student Paper Competition Submission

Shashi Pingolia, Mengwei Li, Tianyi Yang, Matthew A. Lanham
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
shashi@purdue.edu; li4003@purdue.edu; yang1193@purdue.edu; lanhamm@purdue.edu

ABSTRACT

We have researched and developed a life insurance bundling recommendation system that identifies among current home or auto insurance policy holders who is most likely to add a life insurance policy product to their existing plans as well as when to optimally recommended a life product to the customer. The motivation for this study is that insurance product bundling is a common practice in this industry. However, the implementation process of matching customers to the right products is not widely known and likely could be improved using analytical frameworks found in other domains. Generally, the life insurance business does not have integrated predictive analytics that can recommend and price policies in the same way as other insurance areas. For example, the property and casualty industry often utilize a combination of generalized linear models, credibility techniques, and credit scoring models as part of its modeling techniques for driving business decisions (Abrokwah, 2016). However, we posit that an empirically validated methodological design for the cross-product bundling recommendation process in the insurance industry is an area that necessitates deeper analytical investigation.

In collaboration with a major insurance company, we develop and deploy a recommendation engine that uses current policy holder information as features into an ensemble of predictive models to identify when to offer a life policy (single premium, term, or whole life) bundle recommendation that is mostly likely to be purchased. Our solution has provided the insurance company a more efficient, analytically driven, and scalable approach to sell additional products that their customers really want and increase their business revenue. We believe our methodology connects the recommendation system literature to the insurance industry and can be easily adapted by practitioners in this field.

Keywords: Insurance, Recommendation Systems, Predictive Modeling, Analytics, Bundling

INTRODUCTION

It is a common practice to purchase one or more type of insurance policy from the same company. This practice is commonly known as bundling. Insurance bundling is a win-win trade practice for insurance companies and for customers. For customers, it saves money and effort as insurance companies offer multipolicy discounts or multiple line discounts as a means to save on overall insurance bills of their loyal customers. For insurance companies, it is a source of additional revenue from the same customer. Research has also indicated that customers who chooses the benefits of insurance bundling stay more loyal to their insurance companies (Travers 2020).

The bundling practice is beneficial in many cases, but it is not perfect. One of the drawbacks is the tendency to buy unnecessary coverage to get the best savings. For example, a bundle might include an extra $15,000 worth of auto coverage which a customer does not need. It is also possible that separate policies might provide lower rates and other benefits such as auto accidents and ticket forgiveness that are unavailable with bundled policy (Travers 2020).

There are more than 600 companies in the United States alone offering insurance policies (Govindarajula 2019). Generally, insurance companies perform an underwriting process to assess critical factors such as risk of life insurance applicants, policy pricing, etc. The process they follow is arduous, time consuming and involves extensive resources. Sometimes the prolonged time taken in the process causes customers to lose interest and the firm to lose business (Mike Betty 2010).

With the advent of data analytics, it is possible to complete necessary processes much faster and in an efficient manner. This research is intended to do the same for our industry collaborator, a leading mid-west multi-line insurance carrier, offering auto, home, and life insurance products. However, we posit our approach could be generalized to other similar companies. Life insurance is an important segment in our partner's business. However, compared to their auto, home, and farm insurance businesses, which are mainly property and casualty (P&C) insurance, the potential for revenue growth in their life insurance line is seemingly low hanging fruit. We show in our methodological design how one could take advantage of information about their existing customer base from the P&C product line to recommend life insurance products to these existing customers. Our proposed recommendation system could also assist buyers in navigating through customized types of life insurance products. By identifying their current basic policy details and supplying demographic information, the system could recommend the best product for them.

Importance of timing: approximately 10% of the auto insurances are result in some kind of claims whereas the rate is less than 1% in life insurance sector. The low frequency of claims in life insurance as compared to auto insurance presents challenges in developing predictive models. In our case, it intensifies the problem even further because the data is related to only one organization. In life insurance, the insurance span ranges from a year to 20 years whereas in auto insurance, its generally for six months to one year. By predicting the lag between purchasing home or auto policy and buying a life insurance policy, we can identify the right time to approach an existing customer for their life insurance need. Our research helps in mitigating the time and quantity factors and provides a novel solution to improve the existing customer base.

Thus, our business problem is focused on prototyping the ability to accurately recommend a life product policy and bundle it with an existing other non-life-type of insurance policy the customer already possesses, at a time when they are most likely to bundle.

The remaining paper is organized as follows. In the Literature Review we describe related studies focused on recommending and bundling products. In the Data section we describe the data available to us that helped shape our methodological design and models developed. In the Results section we describe the expected empirical recommendation performance of our solution, and in the Conclusions section we provide interesting considerations and future ideas that practitioners might consider when developing their own bundling recommendation system.

LITERATURE REVIEW

A recommender system (RS) is a system that collects information on the preferences for a group of items from the users and facilitates them to make decisions from among the existing alternatives. A tremendous number of recommendation systems are evolving today in parallel with the growth of information in web (Selva Rani and Kumar 2018).
As one of the financial industries, the insurance industry is now facing a vast market and significant growth opportunities. The insurance company will generate many data transactions each day, thus forming a huge database. Recommending insurance products for customers accurately and efficiently can help to improve the competitiveness of the insurance company. Data mining technologies such as association rules have been applied to the recommendation of insurance products (Xu, Wang et al. 2014). However, when large policyholders' data are processed with an associate rule algorithm, it not only requires higher cost of time and space, but also can lead to recommendation rules that lack accuracy and differentiation (Xu, Wang et al. (2014).

An extensive review of the literature suggests there are not many documented instances of recommender systems for the insurance domain. The intelligence recommendation framework for insurance products developed by Xu, Wang et al. (2014) segments customers via cluster analysis and then uses neural networks and the Apriori algorithm to build the product recommendation model. The authors emphasized that using insurance premiums and claims for developing consumer segmentation model can facilitates in reflecting clear value of policyholders. The authors also experimented on the impact that unbalanced data preprocessing can have on the recommendation engine and concludes that the customer segmentation makes the insurance product recommendation more targeted (Xu, Wang et al. 2014)

A car insurance recommendation system framework developed by Lesage, Deaconu et al. (2020) combined the XGBoost algorithm and Apriori algorithm to identify what customer should be contacted and which coverage to recommend, respectively. Their pilot phase testing included 150 recommendations providing evidence that it outperformed standard up-selling approaches at the firm. The framework mentions popular types of recommendation system algorithms such as collaborative filtering, content-based filtering, and hybrid filtering and they elaborate on why most of these popular approaches are more suited for online markets such as entertainment, education, etc. and would not work well for insurance cover recommendations.

Their justifications come from two primary aspects. First, recommending an inappropriate insurance policy could significantly damage the insurer's trustworthiness. Lastly, selecting an insurance policy depends on many complex constraints such as a customer's age, vehicle characteristics, no claim bonuses, and past driving records, etc. These constraints would be challenging to for those algorithms to capture the true signal of what policy is best for customer and the firm (Lesage, Deaconu et al.2020).

Qazi, Tollas et al. (2020) developed a life insurance recommendation system using a deep learning-based approach and general-purpose Bayesian network. Their initial iteration was focused on three states and used two models, one for auto and one for property insurance. Their second iteration, extended the scope to 19 states and which they found was more scalable. The system was trained on data from the firm's customer base, but the models were based solely on external data that can be used for different type of industries. The vast amount of data they used made the computations practically infeasible with BayesiaLab and motivated them to design a mor scalable approach to BN structure.

Table 1 provides a summary of the recent insurance recommendations and how our study compares to those. On the basis of aforesaid literature review and motivation from a leading insurance carrier, it is evident that further study is required to prototype a bundling recommendation engine in the insurance domain.

Table 1: Literature review

Studies	**Domain**	**Methods Used**	**Recommendation Performance**
(Xu, Wang et al. 2014)	Insurance	K means Clustering, Neural Network, Multiple Logistic regression	K means Clustering for more targeted product recommendation
(Lesage, Deaconu et al. 2020)	Car insurance	Gradient Boosting, Random Forest,	Combination of XGBoost and the Apriori algorithm
(Qazi, Tollas et al. 2020)	Life Insurance	Neural Network, Bayesian models	Bayesian model
Our study	Life Insurance Bundling	Linear Regression, Decision Tree, Neural Network, Random Forest, Gradient Boosting	Linear Regression

Recommendation systems that rely either on the classic Apriori algorithm or on Bayesian Networks (BNs) and can be improved to respond as per the actual situation which can differ customer to customer. BNs have smaller memory requirements and allow for faster predictions but require a learning phase that can take a significant amount of time.

Most of the current recommendation models are using neural networks (or deep learning variants), multiple logistic regression, decision trees, support vector classifiers, etc. and there is an immense scope of exploring a standard procedure or method that can help in achieving overall optimization of the parameter of the model. As customers become more technologically savvy, their preferences evolve. The customer is also willing to share additional personal data to get highly personalized services and holistic experiences. An example of this is the use of Telematics data to improve car insurance policy pricing (Simon Jones, Ho-Min Liu et al. 2019). In this context, insurers must rethink their strategies if they would like to stay competitive in a highly saturated and competitive market. Updating their portfolio to provide across segment products, more relevant offering and bundling products has the potential to significantly increase revenue and increase customer loyalty. For example, gig economy workers generally do not have access to employer sponsored insurance benefits and therefore have coverage gaps that insurance firms can fill by bundling different insurance needs (Capgemini 2019).

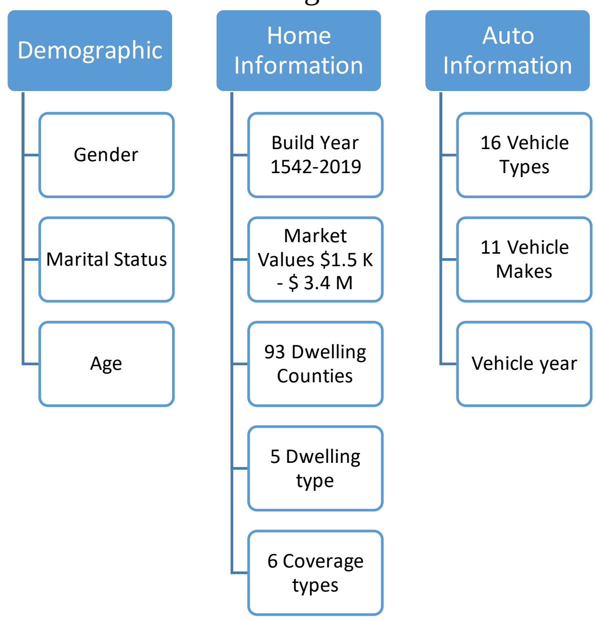
Some of the other important research areas include improving the exploitability of recommendations, i.e., why the specific recommendation was chosen over other options; integrating specific life events such as purchasing a house, purchasing a car, etc., in the recommendation systems so that the insurer need not to adjust their cover frequently; also changing the assumptions to suit customer's requirements.

Our research contribution provides an empirically justified methodological design that recommends who is likely to want to bundle life insurance with their existing auto or home policies, what product is best for them, and when to offer the bundle as timing can impact the buy decision.

DATA

Data was provided by a leading midwestern insurance company carrying auto, home, farm, and life insurance products. Data dictionary tables are provided in the Appendix of this paper for the reader. The data was policy oriented, and we decided to treat the different policy information separately for data modelling. The details of three categories as depicted in Figure 1 are discussed below.

Figure 1: Categories of features used for modeling



We had customer related demographics, such as gender and marital status. The more detailed measurement features are policy related. There are two main categories here which is the auto insurance and the home insurance. Based on the information from different types of insurance policies, we predicted a third insurance type: life insurance, and the best time to promote the correct type of insurance to target customer. Some of the variables we used from the auto insurance are classification codes for the registered vehicle, the registration county. Some of the variables from the home insurance are the dwelling address county and type, market value, replacement value, home coverage type and amount. There was also a derived variable which is the time gap between policy effective dates, which was used to predict the amount of time it takes for a specific individual to choose to purchase a life policy after having an auto or home policy.

Imbalanced Distribution of Project Statuses- Our dataset contains three policy types which are either “Term”, “Whole Life” or “Single Premium”. The pie chart below illustrates the policy type distribution. We can clearly identify data imbalance across the three classes. Single Premium only takes up 1.48%

Data Preparation

The data set had missing values when joining when someone has a lone auto, lone home, or both. After verifying with our industry collaborator, we discovered that our datasets are policy oriented. Meaning that we needed to treat auto/home policies separately given that they have completely unrelated attributes.

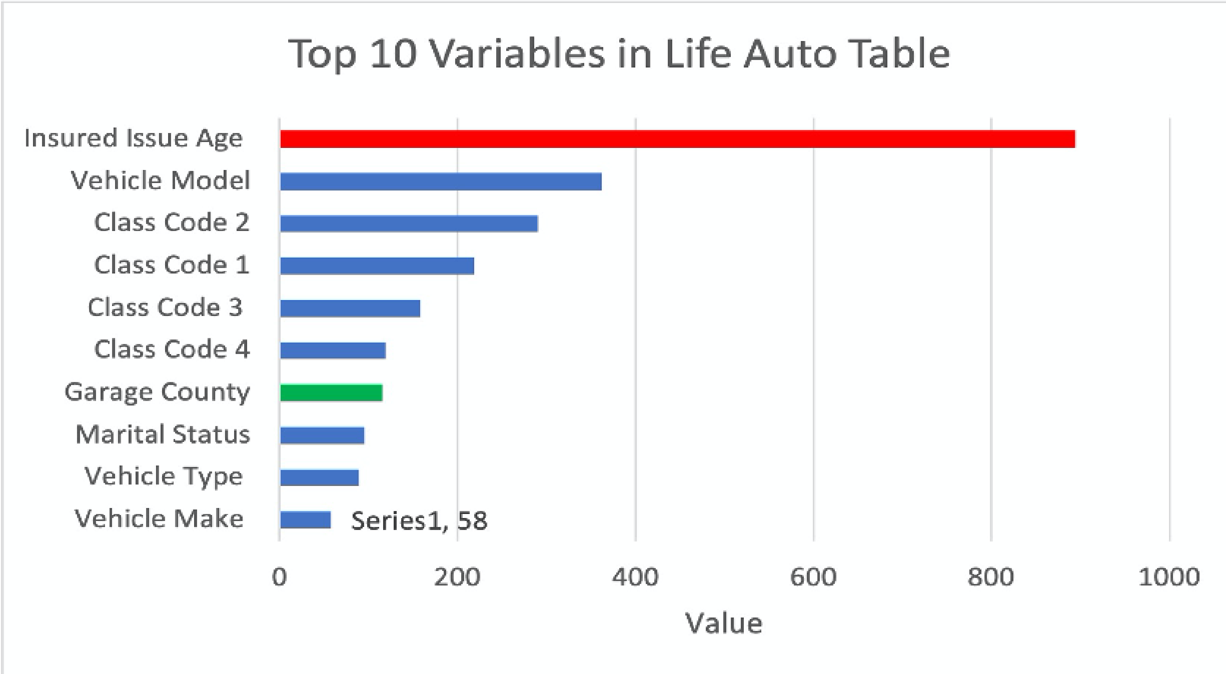
Our business objective was to identify the suitable policy type and the right time to approach a customer. The inception date of each different policy is found to be an attribute that can be used to consider the right time to promote a life policy. The time span between two policy active dates was used as a target variable to predict when to offer the bundled product.

Some fields (e.g., in force and inactive) identified if a policy had expired or not. There were different reasons behind the status. It could be due to cancellations at will, expiration, or upgrading to a new face amount so that the original policy was cancelled. This field was not considered as a useful feature when building the model.

An intriguing aspect of our data preparation process that other firms will likely be tasked with considering is that multiple policies can be held be a single individual. Thus, a traditional left outer join would lead to multiple missing records for those individuals having only a home policy or auto policy. Thus, we joined the life-auto and life-home in two different main tables instead of combining all the tables into one, because the home and auto insurances are totally different types of policies. Additionally, partial columns with same name from different tables are the same while others are not. We need to understand each table and the relationship behind each column before we merge the table. We did a sanity check and examined whether the data for the same person are the same. If they are, we choose to ignore this information from either one of tables or rename the columns to differentiate them when joining.

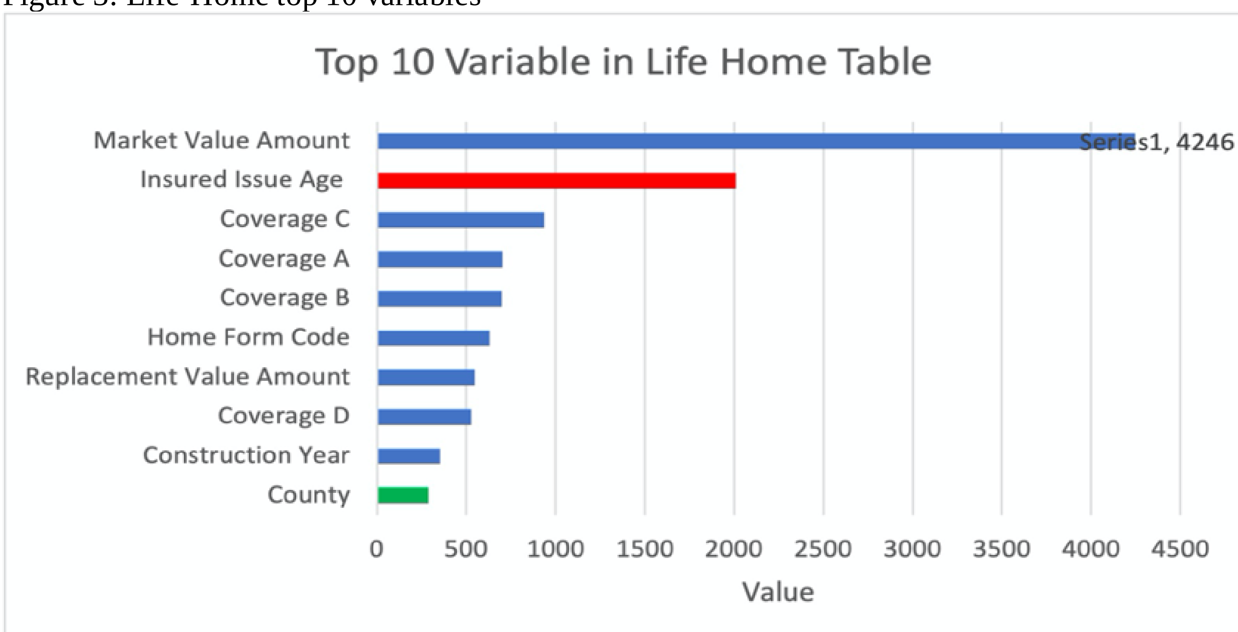
We predicted the more important variables in our data analysis and decided to consider the top 10 variables for model building. It is also worth mentioning that insured issues age and garage county emerged as the important variables in both the datasets (i.e., Life-Auto and Life-Home). The top 10 variables for both the data sets are shown in Figure 2.

Figure 2: Life-Auto top 10 variables



The market value amount and age were the top variables as depicted in Figure 3.

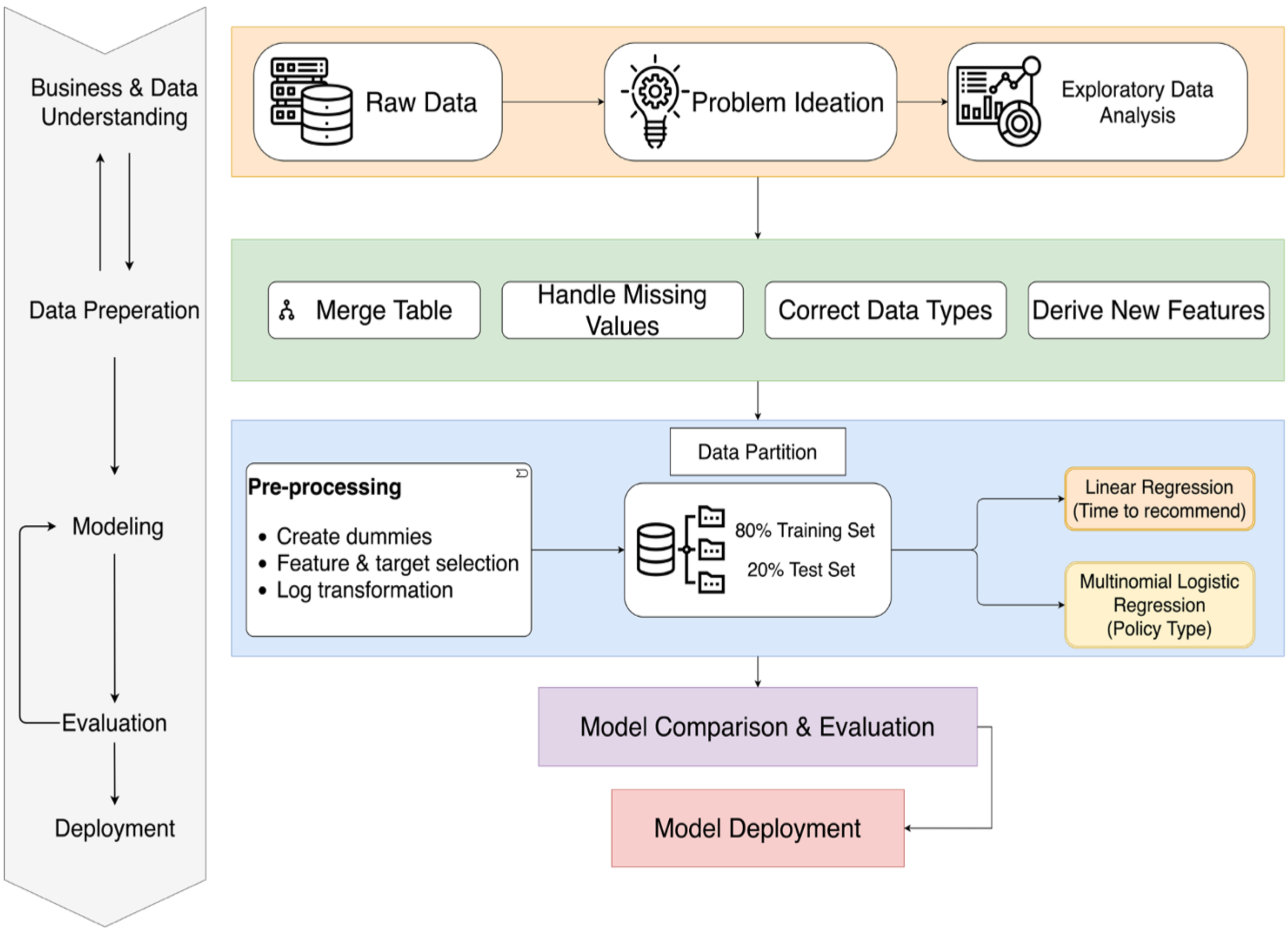
Figure 3: Life-Home top 10 variables



METHODOLOGY

Our methodological design for this recommendation system encompasses a three-tiered design. First, we posit that the firm would want to identify if a current customer would likely purchase a life product given their current auto or home products. Secondly, we could use a multi-classification approach to identify the most likely product for them given the available alternatives. These two models together would ensure we did not recommend a life product to someone who really did not want one. As suggested by (Lesage, Deaconu et al. 2020), recommending an inappropriate insurance policy could significantly damage the insurer's trustworthiness. The third tier of our recommendation solution would be to identify when to offer the bundled product. Timing is everything in business and there is no exception here. This three-tiered design is depicted in Figure 4.

Figure 4: Life Insurance Recommendation Bundling System



Binary Classification

The purpose of binary classification model would allow the insurance company to estimate the propensity a customer would have to purchase a new product given their current products (e.g., auto, home, or both). To do this effectively would require observations of customers who currently have life insurance as well as observations of those that do not. In this study, we were provided only observations of those having life insurance thus our bundling recommendation prototype does currently incorporate this when generating recommendations.

Multi-Classification

Once a customer is identified as a likely "bundler" via the binary classification model the company would want to make sure they offer them the right product among a set of competing products. Multi-classification models are widely used today in retail for similar problems. For example, a merchant trying to estimate the probability a customer would purchase a brand of ketchup among a possible line up on the shelf (e.g., Heinz, Hunts, Store brand). The idea is our model would estimate the probability of purchase for each product for each customer and categorize the testing outcome based on the largest probability type, then the firm would offer them the most probable product.

The models investigated in our study include logistic regression, regression trees and XGBoost. We identify which inputs are most important and should be used in regression. The use of min-max normalization was used to ensure that there is no bias towards larger values in the entire data set. These models are integrated with the recommendation system as a combination of binary model providing probability to purchase, multi-classification model tell the probability for each insurance product type and survival model to identify the best time buy for existing auto or home customer to purchase the life insurance. These models do however have their own set of limitations: due to the restricted number of variables available in the data set. Hence, we have used Random Forest in our final solution as it helps predicts the best type of the life insurance to customer with most accuracy in comparison to all others.

Time-to-Event Prediction

In the model of time prediction, we set up the response variable as the lifespan between the time existing customer purchase the auto or home and when they purchase the life insurance. This figure is vital important for the insurance company to map out the best time to recommend the life insurance to existing customers.

We first check if the response variable is close to normality by density plot. Since the response variable is not normal distributed, we take the log of this continuous variable. However, there were negative infinity values which indicates a result of storing either a large number or a result of division by zero. We decided to add one to the response variable y and then take the log on (Y+1). Thus, our target variable transformed to be log(1+Y). We ran model comparisons on different models, tried both linear regression and random forest for the predictions. For linear regression, we performed stepwise to do variable selection so that we finalized the inputs variables on both datasets. When comparing the results of the linear regression and random forest models, we found that the random forest gave us a comparatively higher adjusted r squared. Additionally, we also checked another metric average square error. Since the average squared error for random forest is smaller, we finally decided to use random forest as our model to do the prediction for lifespan. Lastly, we exponential the log(1+Y) and calculated out the specific months for recommending. Additionally, we also printed out the importance of the variables by comparing value mean.

RESULTS

The multi-classification models are evaluated based on their Accuracy, Precision and Recall scores. The models and their evaluation metrics can be seen below in Table 2. Multinomial logistic Regression failed to predict any single premium types. XGBoost model predicted some single premium types but none of them actually belongs to single premium type. The Random Forest classification model without using the class imbalance treatment outperforms the multinomial logistic regression model and XGBoost model. Precision rates for the 'term' policy type are high across all models. This implies that the life policies are most likely to be predicted as term policy, then whole life, then single premium.

Table 2: Multiclassification model performance

Classification Algorithm	Class Weight	Class	Accuracy	Precision	Recall
Multinomial Logistic Regression	None	Term Whole Life Single Premium	0.73	0.74 0.60 N/A	0.95 0.21 0

Random Forest Classifier	None	Term Whole Life Single Premium	0.73	0.75 0.58 0.26	0.94 0.22 0.05
Random Forest Classifier	Balanced	Term Whole Life Single Premium	0.61	0.80 0.42 0.05	0.69 0.41 0.39
XGBoost Classifier	None	Term Whole Life Single Premium	0.68	0.71 0.36 0	0.92 0.12 0.39

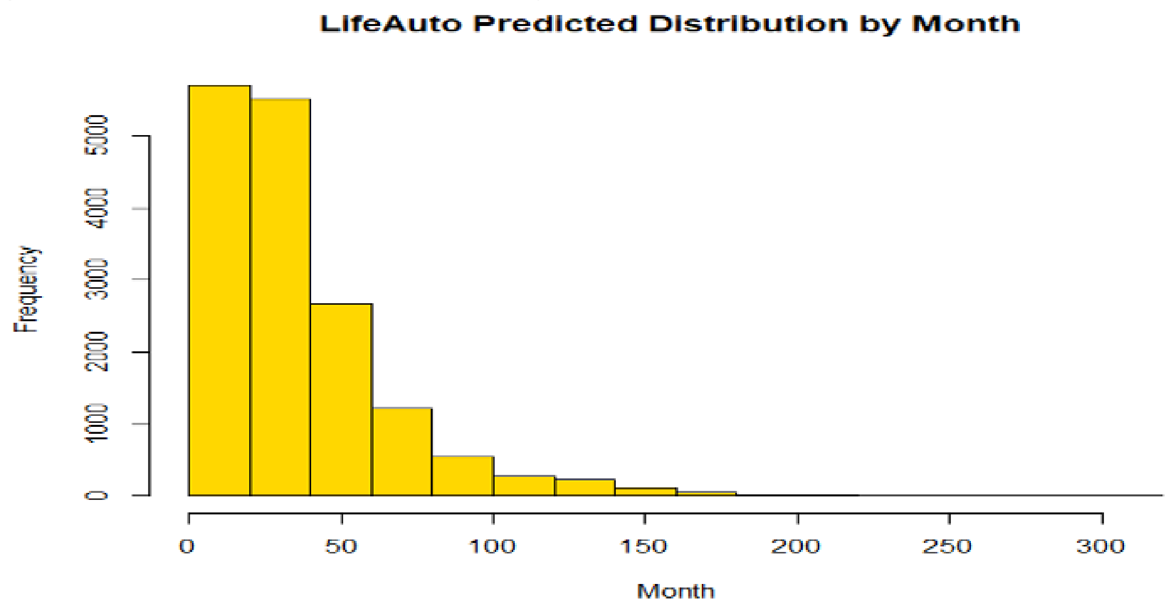
The random forest classifier's base model performs relatively the best among all models. Table 3 is the confusion matrix on the test set. Our model used 500 trees, with 43 variables tried at each split.

Table 3: Confusion matrix of random forest classifier

		Predicted		
		Term	Whole Life	Single Premium
Actual	Term	8577	508	7
	Whole Life	2826	819	19
	Single Premium	86	86	9

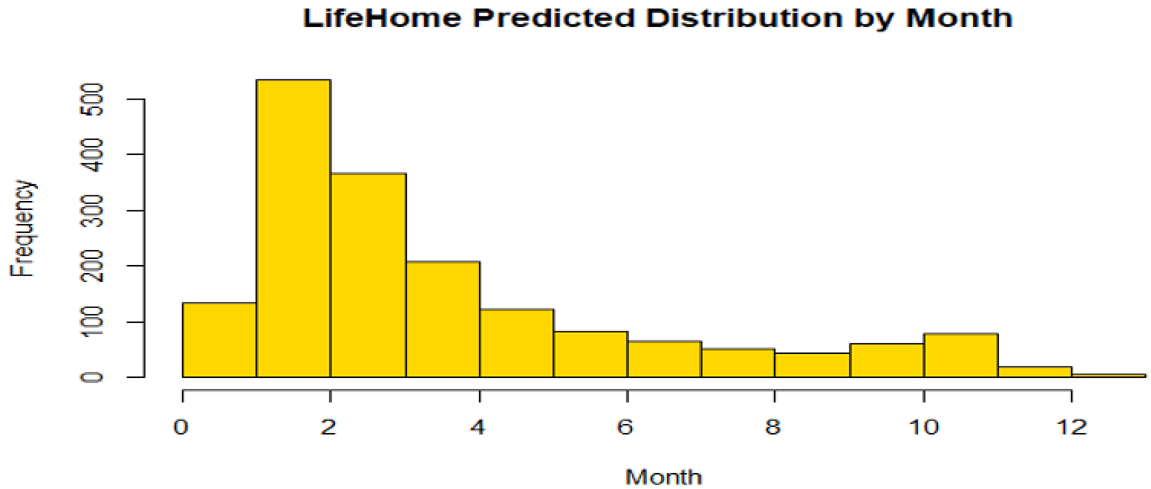
For existing auto insurance customers, the best time to recommend life insurance is within 16 to 55 months later after they purchased auto insurance as shown in Figure 5.

Figure 5: Life-Auto predicted distribution by month



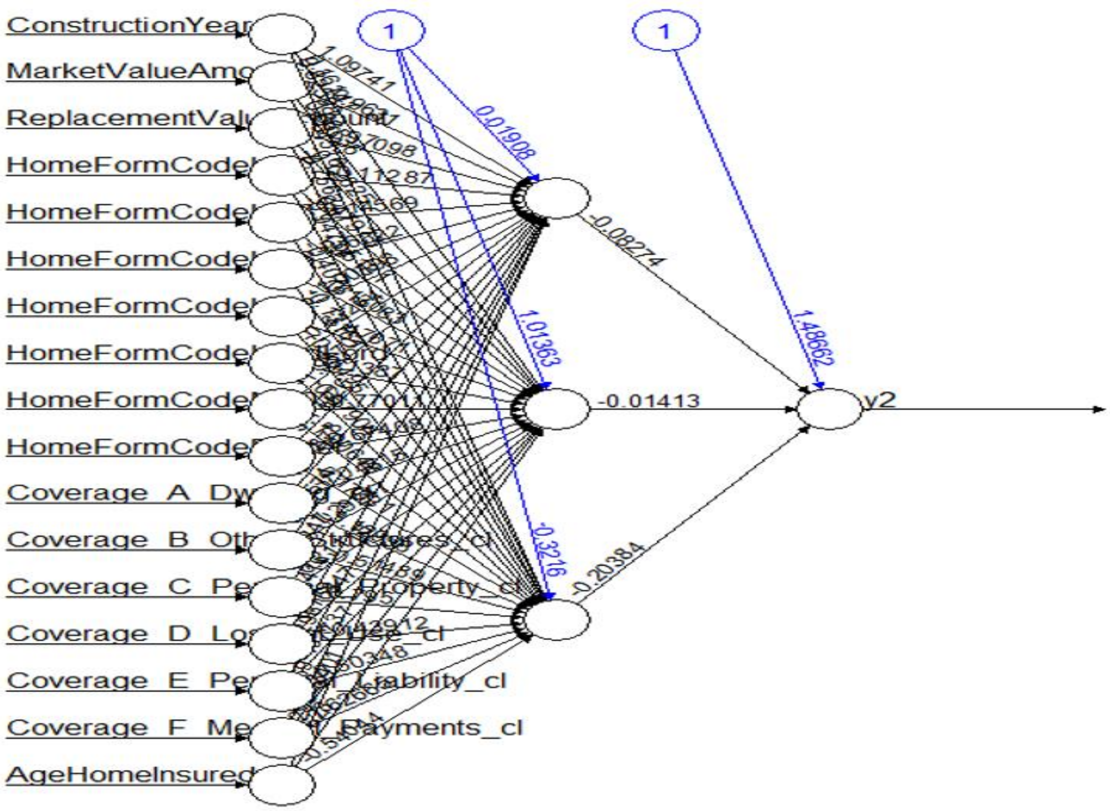
The best time to recommend life insurance to an existing home insurance policy holder is within 1 to 5 months of buying the home insurance policy as depicted in Figure 6. This random forest model was also tuned to 500 ensembled trees, but with 5 variables tried at each split.

Figure 6: Life-Home predicted distribution by month



We experimented 2-layer and 3-layer neural network analysis and the 3-layer model provided better performance. We expected that the ability of the neural network to model non-linear relationship between the internal and external variables would outperform the accuracy of other linear models. However, the accuracy was quite low so we moved on to experiment with other basic models. A depiction of the network is shown in Figure 7.

Figure 7: Neural network model



As our regression problem is about predicting a continuous variable (when to offer the life policy). To evaluate how good our regression model is, we use cross-validation to check the RMSE, R-squared, MAE as our metrics. RMSE tells us how concentrated the data is around the line of best fit, lower values of RMSE indicate better fit. As shown below, we can see the train data set performs better than the test dataset since it has smaller RMSE.

R-squared determines the proportion of variance in the dependent variable (y-lifespan) that can be explained by the independent variable (inputs). In Table 4, R-squared of 37% reveals that 37% of the data fit the regression model for train dataset and 9% of the data fit the regression model for test dataset. Based on MAE, I can certainly interpret that the average difference between the predicted and the actual lifespan. The average difference for train dataset is 1.04 and 1.28 for test dataset. Thus, train dataset fit better.

Table 4: Life Auto performance of predicting when to offer

Life- Auto Table	RMSE	R-squared	MAE
Train	1.31	0.37	1.04
Test	1.59	0.09	1.28

R-squared determines the proportion of variance in the dependent variable (y-lifespan) that can be explained by the independent variable (inputs). In Table 5, the R-squared of 65% reveals that 65% of the data fit the regression model for train dataset and 57% of the data fit the regression model for test dataset. Based on MAE, the average difference between the predicted and the actual lifespan can be calculated. The average difference for train dataset is 0.34 and 0.37 for test dataset. Thus, train dataset fit better.

Table 5: Life Home performance of predicting when to offer

Life-Home Table	RMSE	R-squared	MAE
Train	0.46	0.65	0.34
Test	0.49	0.57	0.37

CONCLUSIONS

Our solution has provided the insurance company a more efficient, analytically driven, approach to recommend and sell additional products that their customers really want and increase their business revenue. We believe our methodology connects the recommendation system literature to the insurance industry and can be easily adapted by practitioners in this field. It is expected that our recommendation system will help our client in developing business strategy to better target customers for life insurance products, and identify the best time to recommend a life policy to existing P&C policy holders.

Our results conclude that random forest performed better than a neural network based on least average squared errors. Our model can identify the correct life insurance type with an overall accuracy of around 70% on all three policy types, which is significant.

In future experiments, a unified model can be developed that unifies the variable in auto and home insurance datasets instead of performing predictive modelling separately. At present our recommendation system focuses on ‘cross selling or bundling of policies’. The future models can focus on ‘the best action to be performed next’ for any customer at any time.

In future, further segregation can be done on the basis of Home types like home owner, renter etc. Similarly, age groups can be further regrouped in smaller categories for both- Home and Auto insurance

Lastly, another research could explore cluster analysis as Xu, W., et.al. (2014) for customer segmentation which can reduce the association rules analysis and can makes insurance products recommendation more targeted. For unbalanced data processing, bagging and price sensitive functions are found to be effective and can be used for more accuracy.

REFERENCES

Capgemini (2019). EFMA World Insurance Report. https://worldinsurancereport.com/.

Govindarajula, S. G. K. (2019). Classifying Risk in Life Insurance using Predictive Analytics. Midwest SAS Users Group (MWSUG) 2019, SAS.

Lesage, L., et al. (2020). "A recommendation system for car insurance." European Actuarial Journal **10**(2): 377-398.

Qazi, M., et al. (2020). "Designing and deploying insurance recommender systems using machine learning." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **10**(4): e1363.

Selva Rani, B. and A. Kumar (2018). "Recommendation Systems: A Beginner's perspective." International Journal of Pure and Applied Mathematics **119**(7): 59-65.

Simon Jones, et al. (2019). IoT in the Insurance Industry: Using Telematics Data to Strategically Manage Risks and Price Competitively. Purdue. 2019 INFORMS Business Analytics Conference, INFORMS**:** 1.

Travers, J. (2020, May 14, 2020). "How to Bundle Home and Auto Insurance." Retrieved 2/1/2021, 2021, from https://www.usnews.com/360-reviews/homeowners-insurance/how-to-bundle-home-and-auto-insurance.

Xu, W., et al. (2014). "A novel intelligence recommendation model for insurance products with consumer segmentation." Journal of Systems Science and Information **2**(1): 16-28.

APPENDIX

Below are data dictionaries of the data used in our study.

Table 1: HomePolicies_New

Variable	Type	Description
PnCPolicyID	Numeric	PnC policy ID
AssociatedLifePolicyID	Numeric	Associated life policy ID
PolicyStatus	Binary	Active, Inactive
InceptionDate	DateTime	date on inception of Home policy
PolicyEndDate	DateTime	End date of Home policies
DwellingAddressCounty	VarChar	County details of Indiana
ConstructionYear	Numeric	Construction year details of Home
MarketValueAmount	Numeric	Market value of home
ReplacementValueAmount	Numeric	Replacement value of home
HomeFormCode	VarChar	Different type of Home Form code
DwellingType	VarChar	Details of dwelling type

Table 2: AutoCoverage

Variable	Type	Description
PnCPolicyID	Numeric	PnC policy ID
AssociatedLifePolicyID	Numeric	Associated life policy ID
Coverage Type	VarChar	Details of different coverage type
LimitAmount	Numeric	Policy coverage amount
MaximumLimitAmount	Numeric	Maximum coverage of policy
DeductableAmount	Numeric	Monthly deductible plan

Table 3: PolicyIndividual_AutoHome

Variable	Type	Description
PnCPolicyID	Numeric	PnC policy ID
AssociatedLifePolicyID	Numeric	Associated life policy ID
PolicyType	Binary	Auto,Home
Relationship	Varchar	Relationship of insurer with policy holder
BirthDate	DateTime	Birthdate of policy holder
Gender	VarChar	Gender of policy holder
MaritalStatus	VarChar	Marital status of policy holder

Table 4: HomeCoverage

Variable	Type	Description
PnCPolicyID	Numeric	PnC policy ID
AssociatedLifePolicyID	Numeric	Associated life policy ID
CoverageType	Varchar	Details of different coverage type
CoverageLimit	Numeric	Policy coverage amount

Table 5: LifePolicies

Variable	Type	Description
LifePolicyID	Numeric	Life policy ID
PolicyStatus	Binary	Active, Inactive
EffectiveDate	DateTime	date on inception of Life policy
ExpirationDate	DateTime	End date of Life policies

LifePolicyType	VarChar	County details of Indiana
InsuredIssueAge	DateTime	Age of policy holder
PolicyFaceAmount	Numeric	Policy coverage amount
IndividualID	Numeric	Policy holder individual ID
Relationship	Varchar	Relationship of insurer with policy holder
BirthDate	DateTime	Birthdate of policy holder
Gender	VarChar	Gender of policy holder
MaritalStatus	VarChar	Marital status of policy holder

Table 6: HomeExtendedCoverage

Variable	Type	Description
PnCPolicyID	Numeric	PnC policy ID
AssociatedLifePolicyID	N7meric	Associated life policy ID
ExtendedCoverageType	Varchar	Details of different extended coverage type
CoverageLimit	Numeric	Policy coverage amount

Table 7: AutoPolicies

Variable	Type	Description
PnCPolicyID	Numeric	PnC policy ID
AssociatedLifePolicyID	Numeric	Associated life policy ID
PolicyStatus	Binary	Active, Inactive
InceptionDate	DateTime	date on inception of Home policy
PolicyEndDate	DateTime	End date of Home policies
ClassCode_1	VarChar	Automobile classification codes
ClassCode_2	VarChar	Automobile classification codes
ClassCode_3	VarChar	Automobile classification codes
ClassCode_4	VarChar	Automobile classification codes
ClassCode_5	VarChar	Automobile classification codes
GarageCounty	VarChar	County details of Indiana
VehicleYear	Numeric	Purchase year of Vehicle
VehicleMake	VarChar	Vehicle Brand
VehicleModel	Numeric	Brand model