KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Optimal mix for selective heterogeneous ensemble learners for churn prediction in the telecommunication industry

Raphael Hanke (0736510)
Koen Barbiers (0726018)
Stefan Hegedis (0606347)

Thesis submitted to obtain the degree:

Master of Science
Information Management

Promotor: Prof. Dr. Vanthienen
Supervisor: Ziboud van Veldhoven

Academic year: 2018/2019

Preface

In the telecommunication industry, churn prediction is an important topic of customer relationship management. Classification algorithms can be combined into a heterogeneous ensemble learner to increase the predictive performance. In this study, 210 combinations of nine common types of classifiers were evaluated in order to discover the optimal mix of base classifiers in a heterogeneous ensemble for churn prediction. The combination of an artificial neural network, a decision tree and gradient boosted trees was identified as the optimal mix. This combination was found to be dominant in ensembles of three classifiers as well as in ensembles of five classifiers. The results of this research confirmed that heterogeneous ensembles can outperform homogeneous ensembles. Besides this, it was found that larger ensembles achieved better scores on average. However, smaller ensembles obtained the best scores. Ensembles that contained top performing base classifiers performed better on average. In the top performing ensembles, the diversity of classifiers plays a role as well.

Leuven, 16/05/2018.

Contents

List of Tables

Chapter 1

Introduction

In many industries, the acquisition of new customers is more expensive than the retention of existing customers (Gallo, 2014). Therefore, customer retention is an important topic of customer relationship management. This holds especially true for the telecommunication industry. Since the competition in the telecom industry is high, telecom companies have a high commitment to prevent churn. A common strategy to prevent churn is offering promotions to customers who are likely to churn. This strategy requires techniques that can predict churners in order to target the right customers.

Machine learning techniques are widely used in the context of customer churn prediction. Each type of classifier has its advantages and disadvantages. Ensembles of different types of classifiers, so called heterogeneous ensembles, can outperform single classifiers as well as homogeneous ensembles which use only one type of classifier.

Most studies on heterogeneous ensemble learners for churn prediction focus on evaluating a single ensemble. In this thesis, the focus is on comparing different ensembles. The aim is to discover the optimal mix of classifiers in a heterogeneous ensemble learner for churn prediction. To achieve this goal, 210 combinations of common types of classifiers are compared. Since four different datasets are used, it is possible to examine which mixes perform the best in general.

The next chapter describes the problem statement and research questions. It is followed by an overview of the relevant literature in chapter three. Chapter four describes the datasets and the preprocessing methods. Chapter five presents the research methodology. The results are discussed in chapter six. Finally, the conclusions of this thesis are presented in chapter seven.

Chapter 2

Problem Statement & Research Questions

Various machine learning algorithms are suitable for churn prediction. Combining different algorithms into an ensemble can improve the predictive performance. To construct such an heterogeneous ensemble two decisions have to be made. First, it is necessary to select base classifiers for the ensemble. Second, an aggregation method needs to be chosen. This thesis focuses on the first decision.

The objective of this thesis is to discover the optimal mix of classifiers in a heterogeneous ensemble learner for churn prediction in the telecom industry. The findings of this thesis give an insight in the characteristics of well-performing ensemble learners. Furthermore, the findings can be used to evaluate existing ensemble selection methods. To achieve the goal of this thesis, the following research questions need to be answered:

1. Do heterogeneous ensembles outperform homogeneous ensembles and single classifiers?
2. What is the influence of the number of base classifiers on the performance of heterogeneous ensembles?
3. Which base classifiers drive the performance of an ensemble?
4. Which pairs of base classifiers are included the most in top performing ensembles?

Chapter 3

Literature Review

Scientific literature on ensemble learning techniques for churn prediction in the telecom industry is extensively available. The literature captures all steps that are involved in the development of complex learning models. The following section will discuss modern research on ensemble techniques for churn prediction in the telecom industry. Since ensemble learners are built up on one or more types of base classifiers, an overview of commonly used classifiers for churn prediction is given first. Secondly, different types of ensembles are discussed. Finally, ensemble selection and aggregation methods are discussed.

Base Classifers

Mahajan, Misra, and Mahajan (2015) in their Review of Data Mining Techniques for Churn Prediction in Telecom analysed 100 articles from 2000 to 2014 to get an overview of the different machine learning techniques that are used for churn prediction in the telecom industry. The authors showed that the most popular methods are decision trees (DT), followed by logistic regression (LR) and artificial neural networks (ANN). Although less frequently, clustering, Naive Bayes (NB), support vector machines (SVM) and k-nearest neighbors (KNN) are used as well to predict churn. Tsai and Lu (2010) and Geeta and Shashi (2012) confirmed these findings by using a similar approach. However, according to Wolpert and Macready (1997), no technique can definitely be proven to be the best in general. The strengths and weaknesses of a classifier always depends on the given data and other related factors, such as the quality and form. Furthermore, each technique is based on unique assumptions or underlies inductive biases, which can bring advantages and disadvantages in the context of churn prediction.

Ensemble Learners

Ensemble learners can be defined as "a set of classifiers whose individual decisions are combined in some way to classify new examples "(Dietterich, 2000, p.1). In various studies ensemble learners outperformed single base classifiers on different performance measures. Ensemble learners demonstrate their strength on large datasets or in cases with complex or nonlinear decision boundaries. The use of ensembles on datasets with too-little data or imbalanced data is beneficial as well (Polikar, 2006).

Homogeneous Ensemble Learners

"Homogeneous ensembles are composed of classifiers of the same type" (Sabzevari, Martínez-Muñoz, & Suárez, 2018, p.1). The most widespread homogeneous ensemble techniques are bagging and boosting. By making use of resampling techniques based on bootstrapping, different training sets are obtained. Subsequently models are iteratively trained with a single type of base classifier and their predictions are combined. Unstable machine learning techniques are the most suitable for bagging, therefore i.e. decision trees are convenient. A minor difference in the resampled data leads to a different model, and thus the ensemble of such models lead to a more diverse learner (Blockeel, 2018).

Vafeiadis, Diamantaras, Sarigiannidis, and Chatzisavvas (2015) discovered improvements through boosting with AdaBoost in ANN, DT and SVM, of which SVM performed the best. Similarly, Buhlman (2012) stated that it is empirically demonstrated that the boosting algorithm AdaBoost is very accurate. It is remarked that the boosting technique requires the tuning of free parameters. Therefore, it can not be applied on NB and LR. Polikar (2006) stressed "the typical consensus, that boosting usually achieves better generalization performances" (p. 39) than bagging. However, the sensitivity to noise and outliers was highlighted. Furthermore, the existence of an universally best ensemble algorithm or combination rule was denied and it was stated that all have proven their effectiveness.

Heterogeneous Ensemble Learners

Heterogeneous models are ensembles which consists of a combination of different base classifiers. Whalen and Pandey (2013) stressed that especially in domains in which disagreement about the optimal base classifier exists, heterogeneous ensemble learners perform stronger than homogeneous ensemble learners. Considering the reviewed literature up to this point suggests an advantage of heterogeneous ensemble learners in the field of churn prediction. This conclusion is consistent with research of i.e. Gilpin and Dunlavy (2009) and Chali, Hasan, and Mojahid (2014). Like many others, their results demonstrated that heterogeneous ensemble learners perform better than homogeneous ensembles, in terms of accuracy through achieved diversity. Besides an increased performance, they are more tolerant to irrelevant attributes in comparison to single base classifiers (Gashler, Giraud-Carrier, & Martinez, 2008). The performance of homogeneous ensembles can be further improved by their inclusion in heterogeneous models (Whalen & Pandey, 2013). This is confirmed by (Gilpin & Dunlavy, 2009) who were able to improve accuracy in their experiments on 4 different dataset by doing so.

Various heterogeneous models for the purpose of churn prediction have been created. Keramati, Aliannejadi, Ahmadian, Mozzafari, and Abbasi (2014) proposed an heterogeneous ensemble learner of ANN, KNN, DT and SVM which outperformed each base classifier in terms of precision and recall. De Caigny, Coussement, and De Bock (2018) focused on the individual problems of decision trees and logistic regressions which led to the development of the Logit Leaf Model (LLM), an ensemble of both. While LR can not cope with interaction effects between variables, decision trees struggle with linear relationships between variables. Trying to predict churn, first a segmentation took place by using the decision tree, consecutively the logistic regression was applied on each cluster. A solution for "situations where the predictor and target variables exhibit complex nonlinear relationships" (p.1) was suggested by Zhang, Qi, Shu, and Cao (2007). Their hybrid classifier that combined KNN and LR improved the classification accuracy and showed superior performance over an ANN+DT(C4.5) hybrid. However, it got outperformed by a single C4.5 decision tree. This demonstrates that ensembles do not always outperform base classifiers. Lee and Lee (2006) introduced an ensemble named SEPI that combines DT(C.5.0), ANN and LR. Huang and Kechadi (2013) combined weighted k-means clustering and a classic inductive rule learning method into an ensemble named FOIL. Outstanding performance was proven in comparison to k-means, DT, LR, PART, SVM, KNN, and OneR and other Hybrid techniques like K-NN+LR, SEPI, for prediction churn.

Ensemble Selection

The discussed studies on heterogeneous ensemble models for churn prediction focus on evaluating a single ensemble. The studies demonstrate that many base classifier combinations are possible when constructing a heterogeneous ensemble learner. This stresses the importance of discovering an optimal mix of base classifiers. The aim of this thesis is to discover this optimal mix by comparing all possible mixes of common types of classifiers.

Given a pool of base classifiers, heterogeneous ensembles can be created by simply aggregating all decisions. However, studies have demonstrated that the selection of the right base classifier subset can result in a comparable or better generalization performance (Liu, Dai, & Liu, 2014). The fact that the performance of an ensemble learner can even decrease if the wrong base classifiers are combined (Whalen & Pandey, 2013) stresses the necessity of a strategic and systematic selection procedure. Tsoumakas, Partalas, and Vlahavas (2009) highlighted, besides the predictive performance, the improvement in computational efficiency which results from a decreased number of models in an ensemble.

Polikar (2006) stressed the importance of diversity within an ensemble learner. A set of base classifiers with unique and indifferent decision boundaries is defined as diverse. Various ensemble selection strategies exist. Cruz, Sabourin, and Cavalcanti (2018) discussed newly developed dynamic selection approaches that create ensembles on the run for each classified instance. Tsoumakas et al. (2009) defined three categories in which existing ensemble selection techniques can be divided: ordering-based, clustering-based and optimisation-based. They found that ensembles perform bad when the diversity is ignored and only the predictive performance of each classifier is taken as selection criteria. Another issue is the determination of the amount of classifiers to select. According to Whalen and Pandey (2013), it is only possible to create an ensemble learner that performs better than the base classifiers if the selected base classifiers are diverse and accurate. Therefore, evaluation measures of ensemble selection methods should be based on these two parameters.

Ensemble Aggregation

After the optimal mix for the heterogeneous ensemble learner is identified, the results of the base learners need to be aggregated. The fusion of selected classifiers is a main issue in ensemble learner design, as it attempts to combine the individual strengths. The different aggregation methods can be divided between non-trainable and trainable approaches (Polikar, 2006).

A simple non-trainable method is majority voting. According to Polikar (2006), there are three types of majority voting: unanimous voting, simple majority voting and weighted voting. All types count the classification of each base classifier as a vote. The first type, unanimous voting, selects a class if it is predicted by all base classifiers. The second type requires that more than half of the base classifiers predict the same class. In the last type, a weight is added to the vote of some classifiers if it is known that these classifiers have more expertise.

Meta learning approaches are widespread as well. Stacking (stacked generalization) is a meta-learning technique that trains a (level-1) classifier on top of the output of a number of (level 0) base classifiers (Whalen & Pandey, 2013). The level 1 classifier learns how to optimally combine the base classifiers. Preliminary work on staked generalizations was undertaken by Ting and Witten (1999), who recommended logistic regressions as level 1 meta classifier as it helps to avoid overfitting and therefore results in superior performance.

Summary

For churn prediction, some base classifiers are more suitable than others. However, none can be clearly identified as the most appropriate. It was discovered that in domains with disagreement about the optimal classifier, heterogeneous ensemble learners outperform homogeneous ensemble learners. The executed literature review underlines the advantage of heterogeneous ensemble models over single base classifiers and homogeneous ensembles for churn prediction. Ensemble learners can outperform the single classifiers by taking advantage of the diversity across classifiers.

Choices need to be made in the process of base classifier selection. The combinations of base classifiers in ensembles for churn prediction in previous studies differ from each other. Neither a clear rule for combinations nor common sense about their construction was discovered. The idea that ensemble selection should not only focus on the combination of highly accurate individual classifiers, but as well on a diversified set of base classifier is widespread. Further improvement can be achieved by the chosen aggregation method. While majority voting convinces through simplicity, meta-learning can be very effective as well, but is more complex and must be applied with caution and respect to the characteristics of the level 1 classifier.

As the interest of high performing models does not diminish, ensemble learners remain a relevant research topic for scientist of different sectors. The recent developments in research on selection procedures highlight the need of guidelines for ensemble construction.

Chapter 4

Data & Preprocessing

4.1 Datasets

Four datasets, named "Duke", "Chile", "UCL" and "Korea", were provided by KU Leuven. Each dataset contains customer data of a telecom company. In general, the features of the datasets capture contract characteristics and customer behavior, such as called minutes, messages and number of complaints. Tables A.1, A.2, A.3, and A.4 in the appendix provide detailed overview. The main characteristics of the datasets can be found in table 4.1.

	Before			After		
Dataset	Observations	Attributes	Churner	Observations	Attributes	Churner
Duke	12499	11	39.31%	12499 (-0%)	12 (+1)	39.31% (+0%)
Chile	7056	46	29.14%	7056 (-0%)	21 (-25)	29.14% (+0%)
UCL	5000	18	14.14%	5000 (-0%)	16 (-2)	14.14% (+0%)
Korea	14490	20	23.11%	13340 (-7.94%)	51 (+31)	23.43% (+0.33%)

Table 4.1: Descriptive statistics of datasets before and after data preprocessing

4.2 Preprocessing

The data quality affects the performance of data mining algorithms (Abbasimehr, Setak, & Tarokh, 2014). By pre-processing the data, the performance of algorithms can be improved significantly. The four datasets were pre-processed by cleaning the data, transforming features and selecting features.

Data Cleaning
The data was cleaned by removing duplicate instances, removing uninteresting features, removing or replacing missing values and treating outliers. The size and other characteristics of the dataset need to be taken into consideration before the application of

any cleaning procedure (De Caigny et al., 2018). Tuples with missing values can be deleted or the missing values can be replaced by using imputation methods. Imputation methods were used when a feature is missing more than 5% of its values. If less than 5% of the values are missing, then the instances were deleted. By doing so, the impact of imputation procedures is limited, as proposed in Verbeke, Dejaeger, Martens, Hur, and Baesens (2012). Depending on the context and the type of the variable, the zero, mean or median imputation procedure was applied (De Caigny et al., 2018). Furthermore, features with more than 30% of missing values were removed (Basiri, Taghiyareh, & Moshiri, 2010). For the detection of Outliers a threshold of three standard deviations from the mean was set. All values that exceeded the threshold were considered as outliers and replaced with the limit (De Caigny et al., 2018).

Only the Korea dataset contains attributes with missing values. Of those six attributes, three have 6.45% missing values, two have 27.85% and one feature has 63.91% missing values. Therefore, for five features the imputation methods were used and one feature was completely removed.

Furthermore, dataset specific issues and anomalies occurred which were solved as follows. Data referring to forced churners, as existent in the Chile dataset, was not considered. ID features were deleted. Moreover, the "UPJONG" attribute in the Korea dataset was deleted, as the cells of this feature contained a non-numerical and non-categorical string in Korean language.

Feature Transformations

Since some algorithms require continuous/numerical variables as input, feature transformations were necessary. By applying one-hot-encoding, categorical features with more than two different values were transformed into binary variables. Furthermore, numeric variables were standardized by applying min-max-scaling, in order to bring them to the same scale with values between 0 and 1. This was essential for building classifiers which are based on distance measures such as KNN or SVM. While maintaining the same distribution, smaller standard deviations limit the effect of outliers.

Class Imbalance

Customer churn datasets are characterized by skewed class distributions, since the amount of customers that churn is very low compared to the amount of non-churners. The given datasets UCL, Chile, Korea and Duke have a churn ratio of 14.14%, 23.11%, 29.14% and 39.31% respectively as stated in table 1. Although the datasets are skewed in the distribution of churners and non-churners, they do not meet the threshold that requires class imbalance correction.

Feature Selection

The extraction of the most influential and meaningful features to build a model is an important part of the data preprocessing (Keramati et al., 2014). Different approaches are used in practice, i.e. Principal Component Analysis (PCA) (De Bock & Poel, 2011), the Partial Decision Tree (PART) algorithm for feature subset selection (Berger, Merkl, & Dittenbach, 2006) or univariate feature selection (Layton, 2015) which is based on correlations between feature and target variable. This research used Recursive Feature Elimination (RFE) to rank the features according to their impact on a the target feature. For Duke and UCL, 11 and 15 predictors were kept, while for Chile and Korea 20 and 50 predictors were kept respectively. The elimination of features is not supposed to improve the performance and to increase computational efficiency.

Chapter 5

Methodology

5.1 Experimental Setup

In the first step, nine base classifiers were created and tuned to achieve optimal performance for each dataset. Simple classifiers as well as homogeneous ensembles were taken under consideration. The simple classifiers were a decision tree, artificial neural network, logistic regression, support vector machine, Naive Bayes and k-nearest neighbors. gradient boosted trees, random forest and AdaBoosted decision tree stumps were the homogeneous ensembles among the candidates. A random forest combines by default 10 decision trees. Gradient boosted trees have by default 100 recursive stages in which a model is fitted on the loss function.

For all nine classifiers, certain hyperparameters were tuned individually for each four datasets by using cross-validated grid search. The hyperparameters were chosen based on background knowledge and subjective opinion of their relevance for the performance of a model. The accuracy measure was used to select the best performing combination of hyperparameters for each classifier. An overview of the base classifiers and their hyperparameters are included in the appendix in table A.17.

For model creation 10-fold cross-validation was used. The main reason for this was the prevention of overfitting and the low amount of instances in UCL and Chile datasets. Furthermore, the whole dataset was shuffled before it was split into training and testing sets. At the end, the results of the 10 folds were averaged.

In the next step, ensembles were build for each dataset. Since the focus of this thesis is not on comparing the performance of different aggregation methods, only majority voting was used. By ignoring more complex aggregation methods as stacking, it was assured that insights into performance improvements were caused by base classifier combination and choice rather than by aggregation strategy. Since majority voting requires an uneven amount of base classifiers, all possible combinations of three and five classifiers were created. This led to 84 ensembles of size three and 126 ensembles of size five.

5.2 Performance Evaluation

Performance measures are indispensable for the evaluation of a model's individual quality and comparisons. Multiple performance metrics exists. Keramati et al. (2014) gave an overview of frequently used performance measures for binary classifiers, including accuracy, misclassification, precision, recall and the F1 score. Other authors make use of measurements such as AUC or TDL (De Bock & Poel, 2011; De Caigny et al., 2018). The choice of the performance measure needs to be aligned to the strategy and purpose of the end user, as each method implies unique characteristics (Verbraken, Verbeke, & Baesens, 2013). In churn prediction the distinction between error types is a major concern. A non-churner classified as a churner (false positive) costs a telecom company less than a churner which can not be identified (false negative) and results to a lost customer. Precision and recall, and therefore F1 score, allow for that distinction. Verbraken et al. (2013) developed with the 'expected maximum profit criterium' a performance measure which is adjusted to the cost issue in churn prediction and focuses on how the profit of a firm is impacted by the performance of the model.

Due to the high costs associated with the false negatives, in this thesis the F1 score was used to evaluate and compare the performance of the created models. False negatives penalize the performance, which allows for a realistic estimation of the model performance, that is in line with the concerns and issues of telecommunication companies in their customer relationship management. Furthermore, for imbalanced data the F1 score is advantageous. Additionally accuracy, recall and AUCroc scores are listed in the tables in the appendix.

Chapter 6

Results

The first goal of this thesis was to discover how the different types of base classifiers perform on their own. Since ensembles are composed out of a number of base classifier, the individual performances of the base classifiers are of interest. Later on, these results are used to determine the relation between the performance of base classifiers and the performance of heterogeneous ensembles. For each dataset table 6.1 ranks the F1 scores of the base classifiers.

Classifier	Duke		UCL		Chile		Korea		AVG	
GB*	1	0.95994	2	0.84236	1	0.89748	4	0.37327	1	0.76826
RF*	5	0.95569	1	0.84277	2	0.86553	2	0.40101	2	0.76625
ANN	4	0.95807	3	0.83655	3	0.83820	5	0.32671	3	0.73988
DT	2	0.95917	4	0.79611	6	0.81300	6	0.26846	4	0.70918
SVM	6	0.93931	5	0.78163	5	0.82700	7	0.17818	5	0.68153
KNN	8	0.92944	6	0.54947	7	0.80088	3	0.40097	6	0.67019
ADA*	3	0.95900	8	0.47457	4	0.83473	8	0.14782	7	0.60403
NB	9	0.86747	7	0.50341	9	0.55283	1	0.40214	8	0.58146
LR	7	0.93144	9	0.20326	8	0.66033	9	0.07932	9	0.46859
AVG	1	0.93995	3	0.64779	2	0.78777	4	0.28643		
SD	1	0.02989	2	0.22626	4	0.10990	3	0.12397		

* Homogeneous ensemble

Table 6.1: Base classifier ranking

On dataset Duke, the base classifiers obtained the highest F1 score on average, followed by Chile, UCL and Korea. The standard-deviation (SD) has to be taken into consideration when interpreting the results. The low SD in Duke denotes a low performance variability between the models. Even the low performing classifiers receive a respectable score. The results of UCL show big differences between top and low performing classifiers. The SD's of Chile and Korea are in between.

LR performed bad on each dataset. The same holds true for NB and KNN, except on the Korea dataset. On the Korea dataset, NB and KNN are top performing base classifiers.

In general, ANN is the best performing single classifier. The results show that the homogeneous ensembles GB and RF outperformed the single classifiers. GB, RF and ADA are all based on decision trees. The high performance of GB and RF is an example of the strength of combining the predictive power of diversified trees. ADA's performance fluctuates compared to GB and RF. ADA obtained a high score on the Duke and Chile dataset and a low score on the UCL and Korea dataset. Based on the results, it can be concluded that in general GB, RF, ANN and DT are more suitable for churn prediction than others. However, the exact suitability differs per dataset.

The first research question was if heterogeneous ensembles can outperform homogeneous ensembles and single classifiers. Table 6.2 ranks for each dataset the scores of the best single classifier, homogeneous ensemble, heterogeneous ensemble of size three and of size five. The results show the advantage of using heterogeneous ensembles over homogeneous ensembles and single classifiers.

Type	Duke		UCL		Chile		Korea		AVG	
Heterogeneous - 3	1	0.96064	1	0.85984	2	0.88772	1	0.43754	1	0.78643
Heterogeneous - 5	3	0.95986	2	0.85832	3	0.88085	2	0.41110	2	0.77753
Homogeneous	2	0.95994	3	0.84277	1	0.89748	4	0.40101	3	0.77530
Single classifier	4	0.95917	4	0.83655	4	0.83820	3	0.40214	4	0.75901
AVG	1	0.95990	3	0.84937	2	0.87606	4	0.41295		
SD	4	0.00060	3	0.01151	1	0.02615	2	0.01701		

Table 6.2: Ensemble type ranking (highest scores)

The highest F1 scores were achieved by heterogeneous ensembles of size three. This holds true for all datasets, except for Chile. On Chile, a homogeneous ensemble performed the best. A single classifer was never the best performing classifier. Heterogeneous ensembles of size five outperformed homogeneous ensembles on the UCL and Korea dataset. On the Duke and Chile dataset homogeneous ensembles scored better. For all datasets except Korea, the score of the best single classifier is lower than the scores of the best ensembles. Based on the results, it can be concluded that heterogeneous ensembles outperform homogeneous ensembles and homogeneous ensembles outperform single classifiers.

The second research question was what the influence is of the number of base classifiers on the performance of heterogeneous ensembles. Based on Table 6.2, it can be concluded that larger heterogeneous ensembles do not outperform smaller heterogeneous ensembles. On all datasets, the best ensembles of size three obtained a higher score than the best ensembles of size five. However, table 6.3 gives a different perspective. For each dataset, table 6.3 ranks the average scores of all single classifiers, homogeneous ensembles, heterogeneous ensembles of size three and of size five. On all datasets, ensembles of size five obtained a higher mean score than ensembles of size three. Based on the results, it can be concluded that adding more base classifiers to an ensemble increases the chance of a higher score. However, smaller ensembles are the best performing ensembles.

Type	Duke		UCL		Chile		Korea		AVG	
Heterogeneous - 5	2	0.95458	1	0.76406	2	0.84954	2	0.30082	1	0.71725
Homogeneous	1	0.95821	3	0.71990	1	0.86591	1	0.30737	2	0.71285
Heterogeneous - 3	3	0.95224	2	0.72562	3	0.83441	3	0.29936	3	0.70290
Single classifier	4	0.93082	4	0.61774	4	0.74871	4	0.27596	4	0.64181
AVG	1	0.94896	3	0.70533	2	0.82464	4	0.29588		
SD	4	0.01234	3	0.06540	1	0.05223	2	0.01373		

Table 6.3: Ensemble type ranking (average scores)

The previous findings, as well as the gap between the average and highest performance, highlight the importance of the optimal combination of classifiers within heterogeneous ensemble learners. In order to determine which base classifiers drive the performance of an ensemble, the partial influence was calculated by taking the arithmetic mean of the F1 score of all classifiers that contain a specific base classifier. Likewise, the indicator for the most impactful pairs was derived. Table 6.4 ranks per dataset the most influential classifier in ensembles of size three and size five as well as the most influential pair in ensembles of size three and size five.

A connection between the individual performances of base classifiers and their influences on ensembles is found. When evaluating the importance of a single classifier within an ensemble, similar results as the base classifier ranking are discovered. For the most influential pair, the results are similar for ensembles of size three and five. The top four most influential pairs are combinations of the four strongest base classifier. This holds for all datasets except of Chile, where two exceptions occur. The decision tree by itself is ranked sixth. However, it is part of the fourth most influential pair for ensembles of size three and part of the third most influential pair for ensembles of size five. Based on this, it can be concluded that ensembles which contain top base classifiers perform better on average. To increase the chance of a high performing ensemble, it is advisable to include the best performing base classifiers. This holds true for ensembles of size three as well as of size five.

According to the literature, the degree of diversity of base classifiers impacts the performance of an ensemble. From the perspective of table 6.4 a-d, it looks like diversity is not important. However, when looking at the top 10 best performing ensembles, diversity plays a role. For each dataset, table 6.5 shows the top 10 best performing ensembles of size three. For each dataset, table 6.6 shows the top 10 best performing ensembles of size five. The best performing ensembles of size three are composed out of one of the four best base classifiers of each dataset. However, the top 10 includes classifiers which were individually ranked lower as well. For example, in Duke the third best ensemble contains LR which is ranked individually seventh. The best performing ensemble of size five on Duke contains LR as well. This suggests that diversity plays a role.

a: Duke

	Single classifier			Classifier pair	
Rank	Base classifier	Ensemble - 3	Ensemble - 5	Ensemble - 3	Ensemble - 5
1	GB	GB	ADA	ADA-GB	ADA-GB
2	DT	ADA	GB	ANN-GB	ADA-DT
3	ADA	DT	DT	ADA-DT	ADA-ANN
4	ANN	ANN	ANN	DT-GB	DT-GB
5	RF	RF	RF	ADA-ANN	ANN-GB
6	SVM	SVM	KNN	ANN-DT	ADA-RF
7	LR	LR	NB	ADA-RF	GB-RF
8	KNN	KNN	SVM	GB-RF	ANN-DT
9	NB	NB	LR	DT-RF	DT-RF

b: UCL

	Single classifier			Classifier pair	
Rank	Base classifier	Ensemble - 3	Ensemble - 5	Ensemble - 3	Ensemble - 5
1	RF	GB	GB	GB-RF	GB-RF
2	GB	RF	RF	ANN-GB	DT-GB
3	ANN	ANN	DT	ANN-RF	DT-RF
4	DT	DT	ANN	DT-RF	ANN-GB
5	SVM	SVM	SVM	DT-GB	ANN-RF
6	KNN	KNN	NB	ANN-DT	ANN-DT
7	NB	NB	KNN	GB-SVM	GB-SVM
8	ADA	ADA	ADA	ANN-SVM	RF-SVM
9	LR	LR	LR	RF-SVM	DT-SVM

c: Chile

	Single classifier			Classifier pair	
Rank	Base classifier	Ensemble - 3	Ensemble - 5	Ensemble - 3	Ensemble - 5
1	GB	GB	GB	GB-RF	GB-RF
2	RF	RF	RF	ANN-GB	ANN-GB
3	ANN	ANN	ANN	ADA-GB	DT-GB
4	ADA	SVM	SVM	DT-GB	ADA-GB
5	SVM	ADA	ADA	GB-SVM	GB-SVM
6	DT	DT	DT	GB-KNN	GB-KNN
7	KNN	KNN	KNN	ANN-RF	ANN-RF
8	LR	LR	LR	ADA-RF	DT-RF
9	NB	NB	NB	RF-SVM	ADA-RF

d: Korea

	Single classifier			Classifier pair	
Rank	Base classifier	Ensemble - 3	Ensemble - 5	Ensemble - 3	Ensemble - 5
1	NB	NB	NB	KNN-GB	KNN-GB
2	RF	RF	KNN	NB-RF	NB-RF
3	KNN	KNN	RF	GB-NB	GB-NB
4	GB	GB	GB	KNN-RF	KNN-RF
5	ANN	ANN	ANN	ANN-NB	ANN-NB
6	DT	DT	DT	GB-RF	GB-RF
7	SVM	SVM	ADA	GB-KNN	GB-KNN
8	ADA	ADA	SVM	DT-NB	ANN-KNN
9	LR	LR	LR	ANN-KNN	ANN-RF

Table 6.4: Most influential classifier/pair per ensemble type (average scores)

Rank	Duke	UCL	Chile	Korea
1	ADA-ANN-GB	ANN-GB-RF	ANN-GB-RF	GB-KNN-NB
2	ANN-DT-GB	GB-RF-SVM	GB-RF-SVM	NB-KNN-RF
3	ADA-GB-LR	ANN-DT-GB	GB-KNN-RF	GB-NB-RF
4	ADA-DT-GB	ANN-RF-SVM	ADA-GB-RF	DT-KNN-RF
5	ADA-ANN-DT	ANN-GB-SVM	DT-GB-RF	ANN-NB-RF
6	ANN-GB-LR	ANN-DT-SVM	ANN-DT-GB	DT-NB-RF
7	ADA-GB-SVM	ANN-DT-RF	ADA-ANN-GB	ANN-KNN-NB
8	ADA-DT-RF	DT-GB-SVM	DT-GB-SVM	ADA-KNN-NB
9	ADA-GB-KNN	GB-NB-RF	ADA-GB-KNN	ADA-NB-RF
10	ADA-GB-NB	GB-KNN-RF	ANN-GB-KNN	RF-NB-SVM

Table 6.5: Top 10 heterogeneous ensembles of size three

Rank	Duke	UCL	Chile	Korea
1	ADA-ANN-DT-GB-LR	ANN-DT-GB-RF-SVM	ADA-ANN-GB-KNN-RF	ANN-GB-KNN-NB-RF
2	ADA-ANN-DT-NB-RF	ANN-DT-GB-NB-SVM	ANN-DT-GB-RF-SVM	DT-GB-KNN-NB-RF
3	ADA-ANN-DT-GB-NB	ANN-GB-NB-RF-SVM	ANN-GB-KNN-RF-SVM	ANN-DT-KNN-NB-RF
4	ADA-ANN-DT-GB-SVM	ANN-DT-GB-NB-RF	ANN-DT-GB-KNN-RF	ADA-GB-KNN-NB-RF
5	ADA-ANN-DT-GB-KNN	ANN-DT-GB-KNN-RF	ADA-ANN-DT-GB-RF	GB-KNN-NB-RF-SVM
6	ADA-ANN-GB-LR-RF	ANN-DT-GB-LR-RF	ADA-ANN-GB-RF-SVM	ADA-ANN-KNN-NB-RF
7	ADA-ANN-DT-LR-RF	ANN-DT-NB-RF-SVM	ADA-DT-GB-RF-SVM	GB-KNN-LR-NB-RF
8	ADA-DT-GB-LR-RF	ADA-ANN-DT-GB-SVM	DT-GB-KNN-RF-SVM	ANN-KNN-NB-RF-SVM
9	ADA-DT-GB-RF-SVM	DT-GB-NB-RF-SVM	ADA-DT-GB-KNN-RF	ANN-KNN-LR-NB-RF
10	ADA-ANN-GB-KNN-LR	ADA-ANN-DT-GB-RF	ADA-ANN-DT-GB-KNN	ANN-DT-GB-KNN-NB

Table 6.6: Top 10 heterogeneous ensembles of size five

The final research question was which pairs of base classifiers are included the most in top performing ensembles. To answer this question, the pairs of base classifiers in the top 10 ensembles of table 6.5 and table 6.6 were counted. Table 6.7a shows the number of occurrences of the pairs that are included in the top 10 ensembles of size three. Only the pairs that occur in all datasets are listed. Since these two pairs don't overlap, it is not possible to combine them into an ensemble of size three. However, when Korea is excluded, there are four more pairs. These pairs are listed in table 6.7b. The exclusion of Korea is justifiable since even the best ensemble obtains a low score on it. The only ensemble of size three that can be constructed out of the six pairs is ANN-DT-GB. For Duke, this is the second best ensemble, for UCL the third and for Chile the sixth. Based on this, it can be concluded that the ensemble ANN-DT-GB is the only ensemble of three classifiers that achieves a top performance on all datasets that allow for a top performance.

	Duke	UCL	Chile	Korea	Total
GB-KNN	1	1	3	1	6
DT-RF	1	1	1	1	4

(a) pairs in all datasets

	Duke	UCL	Chile	Korea	Total
ANN-GB	3	3	4	0	10
DT-GB	2	2	3	0	7
ANN-DT	2	3	1	0	6
GB-SVM	1	3	2	0	6

(b) pairs in all datasets except Korea

Table 6.7: Occurrences of pairs of base classifiers in top 10 ensembles of size three

In ensembles of size five, the combination ANN-DT-GB is dominant as well. Table 6.8a shows the number of occurrences of the pairs of base classifiers that are included in the top 10 ensembles of size five. Only the pairs that occur in all datasets are listed. It is not possible to combine these pairs into an ensemble of five classifiers. However, when Korea is excluded, there are three more pairs. These pairs are listed in table 6.8b. The eighteen pairs in total can be combined into three ensembles of size five. Table 6.9 shows these ensembles and their performance ranks in each dataset. Based on this, it can be concluded that the ensembles ADA-ANN-DT-GB-SVM and ADA-ANN-DT-GB-RF are the only ensembles of five classifiers that achieve a top performance on all datasets that allow for a top performance. Both ensembles include the combination ADA-ANN-DT-GB.

	Duke	UCL	Chile	Korea	Total
GB-RF	3	7	9	5	24
ANN-GB	6	8	7	2	23
DT-GB	6	8	7	2	23
ANN-RF	3	7	6	5	21
ANN-DT	6	8	4	2	20
DT-RF	4	7	6	2	19
ADA-GB	8	2	6	1	17
GB-KNN	2	1	6	6	15
ADA-KNN	8	2	4	1	15
ANN-KNN	2	1	4	6	13
ADA-RF	5	1	5	2	13
GB-SVM	2	5	5	1	13
RF-SVM	1	4	5	2	12
ANN-SVM	1	5	3	1	10
DT-KNN	1	1	4	3	9

(a) pairs in all datasets

	Duke	UCL	Chile	Korea	Total
ADA-DT	8	2	4	0	14
DT-SVM	2	5	3	0	10
ADA-SVM	2	1	2	0	5

(b) paris in all datasets except Korea

Table 6.8: Occurrences of pairs of base classifiers in top 10 ensembles of size five

	Duke	UCL	Chile
ADA-ANN-DT-GB-SVM	4	8	12
ADA-ANN-DT-GB-RF	12	10	5
ADA-ANN-DT-RF-SVM	23	14	41

Table 6.9: Ensembles constructed out of table 6.8 and their rank in each dataset

Chapter 7

Conclusion

The objective of this thesis was to discover the optimal mix of classifiers in a heterogeneous ensemble learner for churn prediction in the telecom industry. The findings of this study give insights into the characteristics of well-performing heterogeneous ensemble learners. To discover the optimal mix, 210 combinations of common classifiers for churn prediction were compared. The pool of base classifiers included a decision tree (DT), artificial neural network (ANN), logistic regression (LR), support vector machine (SVM), Naive Bayes (NB), k-nearest neighbors (KNN), gradient boosted trees (GB), random forest (RF) and AdaBoosted decision tree stumps (ADA). Four different datasets were used to examine which mix performs the best in general.

The first finding of this study was that in general GB, RF, ANN and DT performed better than SVM, KNN, ADA, NB and LR. An insight in the individual performance of base classifiers was needed to determine the relation between the performance of base classifiers and the performance of heterogeneous ensembles. A strong connection was found. Ensembles that contained top performing base classifiers performed better on average. However, in the top performing ensembles the diversity of classifiers played a role as well.

Another finding of this study was that heterogeneous ensembles outperformed homogeneous ensembles, which on their turn outperformed single classifiers. This is in line with the reviewed literature. It shows the advantage of using heterogeneous ensembles for churn prediction. The impact of the number of base classifiers on the performance of ensembles was examined as well. Larger ensembles achieved better scores on average. However, the best performing ensembles were smaller ensembles.

The final finding of this study was that one combination of base classifiers led to an ensemble that obtained a top score on all datasets except one. The classifiers that form this combination are ANN, DT and GB. The combination is found to be dominant in ensembles with three base classifiers, as well as in ensembles with five base classifiers. Therefore, the combination of an artificial neural network, a decision tree and gradient boosted trees can be seen as the optimal mix for a heterogeneous ensemble learner.

The presented study was conducted within certain limits and constraints. Hence, certain assumptions underlie the validity of the results. First, only four churn datasets of the telecommunication industry were used. Second, more types of base classifiers exist, which were not taken into consideration. Third, different hyperparameter settings for the selected base classifiers were possible. Fourth, the data could have been pre-processed differently. Sixth, only ensembles of size three and five were constructed.

The study can be extended to topics other than churn prediction, in order to confirm the findings in multiple contexts. Another possible extension is the evaluation of existing ensemble selection methods. The results of this study can be used to see if the ensembles that are formed by selection methods are top performing ensembles. Examining the influence of different aggregation methods is another topic for further research.

Appendix A

Appendix

A.1 Overview of Features

Feature	Meaning
subID	Customer ID
churnInd	Churner Y/N
longevityMonths	Customer lifetime in months
minLastMonths	Call minutes in the last month
totalMin	Total call minutes
plan	Phone plan
PlanMin	Call minutes included in plan
LastMonthDiff	Difference of call minutes between the last two months
AvgMin	Average call minutes
AvgDiff	Average difference of call minutes between months
promMonth	-

Table A.1: Overview of features - Duke

Feature	Meaning
Area_Code2	-
Intl_Plan	International plan Y/N
Vmail	Voicemail Y/N
Vmail_Message	Voicemail messages
Day_Mins	Total call minutes during the day
Day_Calls	Calls during the day
Day_Charge	Charged for calls during the day
Eve_Mins	Total call minutes during the evening
Eve_Calls	Calls during the evening
Eve_Charge	Charged for calls during the evening
Night_Mins	Total call minutes during the night
Night_Calls	Calls during the night
Night_Charge	Charged for calls during the night
Intl_Mins	Total minutes for international calls
Intl_Calls	International calls
Intl_Charge	Charged for international calls
CustServ_Calls	Customer service calls
Churn	Churner Y/N

Table A.2: Overview of features - UCL

Feature	Meaning
ID	Customer ID
START_DATE	Start of contract
END_DATE	End of contract
ACTIVE_DAYS	Contract duration in days
ACTIVE_WEEKS	Contract duration in weeks
ACTIVE_MONTHS	Contract duration in months
CHURN	Churner Y/N
PREPAID_BEFORE	Prepaid / No prepaid during past
COLLECTIONS	-
PAYMENT_DELAY	Actual delay of payment
ANNUAL_PAY_DELAY	Delays of payments during the year
RECEIPT_DELAYS	Delayed receipts per year
COMPLAINT_2WEEKS	Complaints last 2 weeks
COMPLAINT_3MONTHS	Complaints last 3 months
COMPLAINT_6MONTHS	Complaints last 6 months
COMPLAINT_1WEEK	Complaints last week
COMPLAINT_1MONTH	Complaints last month
ARPU	-
COUNT_OFFNET_CALLS_1WEEK	Off-net incoming calls last week
COUNT_ONNET_CALLS_1WEEK	On-net incoming calls last week
AVG_INC_OFFNET_1MONTH	Average off-net incoming calls last month
AVG_INC_ONNET_1MONTH	Average on-net incoming calls last month
AVG_DATA_3MONTH	Average of bytes last 3 months
COUNT_CONNECTIONS_3MONTH	-
AVG_DATA_1MONTH	Average bytes last month
COUNT_SMS_INC_ONNET_6MONTH	On-net incoming SMS last 6 months
COUNT_SMS_OUT_OFFNET_6MONTH	Off-net outgoing SMS last 6 months
COUNT_SMS_INC_OFFNET_1MONTH	Off-net incoming SMS last month
COUNT_SMS_INC_OFFNET_WKD_1MONTH	Off-net incoming SMS in weekends last month
COUNT_SMS_INC_ONNET_WKD_1MONTH	On-net incoming SMS in weekends last month
COUNT_SMS_OUT_OFFNET_1MONTH	Off-net outgoing SMS last month
COUNT_SMS_OUT_OFFNET_WKD_1MONTH	Off-net outgoing SMS in weekends last month
COUNT_SMS_OUT_ONNET_1MONTH	On-net outgoing SMS last month
COUNT_SMS_OUT_ONNET_WKD_1MONTH	Number of on-net outgoing SMS in weekends last month
AVG_MINUTES_INC_OFFNET_1MONTH	Average off-net incoming call minutes last month
AVG_MINUTES_INC_ONNET_1MONTH	Average on-net incoming call minutes last month
MINUTES_INC_OFNET_WKD_1MONTH	Off-net incoming call minutes in weekends last month
MINUTES_INC_ONNET_WKD_1MONTH	On-net incoming call minutes in weekends last month
AVG_MINUTES_OUT_OFFNET_1MONTH	Average off-net outgoing call minutes last month
AVG_MINUTES_OUT_ONNET_1MONTH	Average on-net outgoing call minutes last month
MINUTES_OUT_OFFNET_WKD_1MONTH	Off-net outgoing call minutes in weekends last month
MINUTES_OUT_ONNET_WKD_1MONTH	On-net outgoing call minutes in weekends last month
MINUTES_INC_ONNET_3MONTH	On-net incoming call minutes last three months
MINUTES_INC_OFFNET_3MONTH	Off-net incoming call minutes last three months

Table A.3: Overview of features - Chile

Feature	Meaning
CustID	Customer ID
CUSTGB	-
REGION	Customer region
PRDCD	Product code
OPNDT	Start date
CLSDT	End date
USEMM	Used call minutes
TOTMM	Total call minutes
REV6	Total customer revenue
AVG6	Average customer revenue per month
CONTACT	Number of customer contacts
MINAP_GIGAN	-
MINAP_AMP	-
UPJONG	-
EMPTOTAL	-
CORPEV	-
JANGCNT	-
CLAIMCNT	Number of claims
PAYMTHD	Payment method
TARGET	Churner Y/N

Table A.4: Overview of features - Korea

A.2 Scores of Base Classifiers

Classifier	Accuracy	Recall	F1 score	ROCauc
ADA	0.96872	0.92935	0.95900	0.96184
ANN	0.96808	0.92721	0.95807	0.96089
DT	0.96904	0.92421	0.95917	0.96117
GB	0.96944	0.93076	0.95994	0.96268
KNN	0.94688	0.89183	0.92944	0.93715
LR	0.94768	0.90571	0.93144	0.94026
NB	0.89975	0.83599	0.86747	0.88857
RF	0.96616	0.92884	0.95569	0.95961
SVM	0.95432	0.90148	0.93931	0.94502

95994

Table A.5: Scores of base classifiers - Duke

Classifier	Accuracy	Recall	F1 score	ROCauc
ADA	0.88360	0.37514	0.47457	0.67102
ANN	0.95720	0.77729	0.83655	0.88184
DT	0.94720	0.73220	0.79611	0.85735
GB	0.95940	0.77244	0.84236	0.88119
KNN	0.90640	0.40798	0.54947	0.69791
LR	0.86680	0.12131	0.20326	0.55552
NB	0.86240	0.49704	0.50341	0.70958
RF	0.96020	0.75817	0.84277	0.87568
SVM	0.94520	0.70206	0.78163	0.84354

Table A.6: Scores of base classifiers - UCL

Classifier	Accuracy	Recall	F1 score	ROCauc
ADA	0.90859	0.79326	0.83473	0.87472
ANN	0.90490	0.84864	0.83820	0.88844
DT	0.89541	0.78105	0.81300	0.86205
GB	0.94119	0.88513	0.89748	0.92462
KNN	0.88549	0.79241	0.80088	0.85838
LR	0.83007	0.56875	0.66033	0.75334
NB	0.76559	0.49886	0.55283	0.68731
RF	0.92574	0.82272	0.86553	0.89548
SVM	0.90079	0.81788	0.82700	0.87646

Table A.7: Scores of base classifiers - Chile

Classifier	Accuracy	Recall	F1 score	ROCauc
ADA	0.77466	0.08387	0.14782	0.53493
ANN	0.77976	0.22865	0.32671	0.58852
DT	0.78501	0.16971	0.26846	0.57131
GB	0.79805	0.25691	0.37327	0.61024
KNN	0.76094	0.34172	0.40097	0.61551
LR	0.76702	0.04300	0.07932	0.51576
NB	0.31507	0.98358	0.40214	0.54700
RF	0.78021	0.31436	0.40101	0.61862
SVM	0.78216	0.10104	0.17818	0.54582

Table A.8: Scores of base classifiers - Korea

A.3 Scores of Ensembles of Size Three

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
1	Y	Y	Y	-	-	-	-	-	-	0.95544	0.90431	0.94088	0.94643
2	Y	Y	-	Y	-	-	-	-	-	0.96896	0.92484	0.95910	0.96122
3	Y	Y	-	-	Y	-	-	-	-	0.95856	0.90436	0.94489	0.94900
4	Y	Y	-	-	-	Y	-	-	-	0.95800	0.89723	0.94375	0.94729
5	Y	Y	-	-	-	-	Y	-	-	0.96840	0.92523	0.95840	0.96083
6	Y	Y	-	-	-	-	-	Y	-	0.96920	0.92421	0.95936	0.96131
7	Y	Y	-	-	-	-	-	-	Y	0.96920	0.92566	0.95943	0.96158
8	Y	-	Y	Y	-	-	-	-	-	0.95536	0.90472	0.94082	0.94644
9	Y	-	Y	-	Y	-	-	-	-	0.95304	0.90123	0.93771	0.94389
10	Y	-	Y	-	-	Y	-	-	-	0.95264	0.89885	0.93708	0.94317
11	Y	-	Y	-	-	-	Y	-	-	0.95496	0.90408	0.94026	0.94599
12	Y	-	Y	-	-	-	-	Y	-	0.95536	0.90430	0.94078	0.94636
13	Y	-	Y	-	-	-	-	-	Y	0.95560	0.90533	0.94115	0.94675
14	Y	-	-	Y	Y	-	-	-	-	0.95800	0.90556	0.94425	0.94874
15	Y	-	-	Y	-	Y	-	-	-	0.95792	0.89827	0.94373	0.94741
16	Y	-	-	Y	-	-	Y	-	-	0.96784	0.92604	0.95773	0.96050
17	Y	-	-	Y	-	-	-	Y	-	0.96904	0.92504	0.95920	0.96133
18	Y	-	-	Y	-	-	-	-	Y	0.96936	0.92730	0.95971	0.96200
19	Y	-	-	-	Y	Y	-	-	-	0.94992	0.88745	0.93297	0.93890
20	Y	-	-	-	Y	-	Y	-	-	0.95688	0.90615	0.94283	0.94793
21	Y	-	-	-	Y	-	-	Y	-	0.95888	0.90497	0.94532	0.94938
22	Y	-	-	-	Y	-	-	-	Y	0.95880	0.90621	0.94529	0.94954
23	Y	-	-	-	-	Y	Y	-	-	0.95752	0.89863	0.94321	0.94713
24	Y	-	-	-	-	Y	-	Y	-	0.95832	0.89784	0.94419	0.94766
25	Y	-	-	-	-	Y	-	-	Y	0.95824	0.89929	0.94418	0.94786
26	Y	-	-	-	-	-	Y	Y	-	0.96864	0.92562	0.95871	0.96109
27	Y	-	-	-	-	-	Y	-	Y	0.96840	0.92770	0.95850	0.96128
28	Y	-	-	-	-	-	-	Y	Y	0.96960	0.92667	0.95997	0.96208
29	-	Y	Y	Y	-	-	-	-	-	0.96888	0.92400	0.95895	0.96100
30	-	Y	Y	-	Y	-	-	-	-	0.96048	0.90619	0.94737	0.95092
31	-	Y	Y	-	-	Y	-	-	-	0.95920	0.89871	0.94537	0.94856
32	-	Y	Y	-	-	-	Y	-	-	0.96840	0.92420	0.95834	0.96064
33	-	Y	Y	-	-	-	-	Y	-	0.96912	0.92380	0.95924	0.96117
34	-	Y	Y	-	-	-	-	-	Y	0.96888	0.92422	0.95895	0.96105
35	-	Y	-	Y	Y	-	-	-	-	0.96840	0.92563	0.95840	0.96090
36	-	Y	-	Y	-	Y	-	-	-	0.96856	0.92421	0.95855	0.96078
37	-	Y	-	Y	-	-	Y	-	-	0.96856	0.92563	0.95860	0.96103
38	-	Y	-	Y	-	-	-	Y	-	0.96944	0.92525	0.95972	0.96169
39	-	Y	-	Y	-	-	-	-	Y	0.96976	0.92771	0.96023	0.96240
40	-	Y	-	-	Y	Y	-	-	-	0.95544	0.89415	0.94032	0.94464
41	-	Y	-	-	Y	-	Y	-	-	0.96688	0.92583	0.95648	0.95968
42	-	Y	-	-	Y	-	-	Y	-	0.96888	0.92401	0.95895	0.96101
43	-	Y	-	-	Y	-	-	-	Y	0.96864	0.92588	0.95874	0.96116
44	-	Y	-	-	-	Y	Y	-	-	0.96872	0.92503	0.95879	0.96106
45	-	Y	-	-	-	Y	-	Y	-	0.96904	0.92442	0.95917	0.96121
46	-	Y	-	-	-	Y	-	-	Y	0.96880	0.92525	0.95891	0.96117
47	-	Y	-	-	-	-	Y	Y	-	0.96928	0.92502	0.95950	0.96151
48	-	Y	-	-	-	-	Y	-	Y	0.96872	0.92606	0.95884	0.96125
49	-	Y	-	-	-	-	-	Y	Y	0.96960	0.92564	0.95992	0.96189
50	-	-	Y	Y	Y	-	-	-	-	0.96000	0.90781	0.94686	0.95080
51	-	-	Y	Y	-	Y	-	-	-	0.95872	0.89933	0.94481	0.94827
52	-	-	Y	Y	-	-	Y	-	-	0.96776	0.92501	0.95756	0.96025
53	-	-	Y	Y	-	-	-	Y	-	0.96872	0.92380	0.95872	0.96084
54	-	-	Y	Y	-	-	-	-	Y	0.96912	0.92627	0.95934	0.96162
55	-	-	Y	-	Y	Y	-	-	-	0.95048	0.88502	0.93346	0.93896
56	-	-	Y	-	Y	-	Y	-	-	0.95880	0.90819	0.94534	0.94988
57	-	-	Y	-	Y	-	-	Y	-	0.96056	0.90639	0.94747	0.95102
58	-	-	Y	-	Y	-	-	-	Y	0.96040	0.90784	0.94737	0.95116
59	-	-	Y	-	-	Y	Y	-	-	0.95856	0.89949	0.94458	0.94816
60	-	-	Y	-	-	Y	-	Y	-	0.95928	0.89890	0.94548	0.94866
61	-	-	Y	-	-	Y	-	-	Y	0.95896	0.89973	0.94511	0.94855
62	-	-	Y	-	-	-	Y	Y	-	0.96872	0.92499	0.95877	0.96104
63	-	-	Y	-	-	-	Y	-	Y	0.96824	0.92645	0.95823	0.96092
64	-	-	Y	-	-	-	-	Y	Y	0.96936	0.92563	0.95961	0.96169
65	-	-	-	Y	Y	Y	-	-	-	0.95528	0.89598	0.94022	0.94482
66	-	-	-	Y	Y	-	Y	-	-	0.96656	0.92666	0.95612	0.95957
67	-	-	-	Y	Y	-	-	Y	-	0.96848	0.92564	0.95851	0.96097
68	-	-	-	Y	Y	-	-	-	Y	0.96856	0.92791	0.95871	0.96145
69	-	-	-	Y	-	Y	Y	-	-	0.96792	0.92521	0.95776	0.96043
70	-	-	-	Y	-	Y	-	Y	-	0.96864	0.92421	0.95865	0.96084
71	-	-	-	Y	-	Y	-	-	Y	0.96888	0.92710	0.95909	0.96157
72	-	-	-	Y	-	-	Y	Y	-	0.96872	0.92582	0.95881	0.96119
73	-	-	-	Y	-	-	Y	-	Y	0.96896	0.92851	0.95924	0.96188
74	-	-	-	Y	-	-	-	Y	Y	0.97008	0.92810	0.96064	0.96273
75	-	-	-	-	Y	Y	Y	-	-	0.95440	0.89698	0.93915	0.94427
76	-	-	-	-	Y	Y	-	Y	-	0.95576	0.89497	0.94076	0.94505
77	-	-	-	-	Y	Y	-	-	Y	0.95592	0.89685	0.94111	0.94552
78	-	-	-	-	Y	-	Y	Y	-	0.96712	0.92641	0.95680	0.95997
79	-	-	-	-	Y	-	Y	-	Y	0.96712	0.92808	0.95689	0.96028
80	-	-	-	-	Y	-	-	Y	Y	0.96920	0.92708	0.95948	0.96183
81	-	-	-	-	-	Y	Y	Y	-	0.96896	0.92562	0.95911	0.96136
82	-	-	-	-	-	Y	Y	-	Y	0.96864	0.92666	0.95876	0.96129
83	-	-	-	-	-	Y	-	Y	Y	0.96920	0.92605	0.95943	0.96164
84	-	-	-	-	-	-	Y	Y	Y	0.96904	0.92668	0.95926	0.96163

Table A.9: Scores of ensembles of size three - Duke

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
1	Y	Y	Y	-	-	-	-	-	-	0.94380	0.63607	0.75985	0.81523
2	Y	Y	-	Y	-	-	-	-	-	0.95200	0.69003	0.79959	0.84232
3	Y	Y	-	-	Y	-	-	-	-	0.91460	0.42026	0.57979	0.70790
4	Y	Y	-	-	-	Y	-	-	-	0.91420	0.47466	0.60813	0.73045
5	Y	Y	-	-	-	-	Y	-	-	0.95600	0.72622	0.82284	0.85994
6	Y	Y	-	-	-	-	-	Y	-	0.89740	0.33357	0.47700	0.66177
7	Y	Y	-	-	-	-	-	-	Y	0.95480	0.72443	0.81775	0.85847
8	Y	-	Y	Y	-	-	-	-	-	0.94920	0.68276	0.78874	0.83775
9	Y	-	Y	-	Y	-	-	-	-	0.91500	0.43107	0.58722	0.71273
10	Y	-	Y	-	-	Y	-	-	-	0.90500	0.43626	0.56258	0.70902
11	Y	-	Y	-	-	-	Y	-	-	0.94920	0.66870	0.78651	0.83200
12	Y	-	Y	-	-	-	-	Y	-	0.89720	0.32695	0.47260	0.65891
13	Y	-	Y	-	-	-	-	-	Y	0.95000	0.67699	0.79011	0.83592
14	Y	-	-	Y	Y	-	-	-	-	0.91540	0.43655	0.59034	0.71511
15	Y	-	-	Y	-	Y	-	-	-	0.90720	0.44409	0.57307	0.71352
16	Y	-	-	Y	-	-	Y	-	-	0.95460	0.70471	0.81202	0.85000
17	Y	-	-	Y	-	-	-	Y	-	0.89940	0.33855	0.48685	0.66507
18	Y	-	-	Y	-	-	-	-	Y	0.95580	0.71318	0.81762	0.85424
19	Y	-	-	-	Y	Y	-	-	-	0.89140	0.32848	0.45841	0.65594
20	Y	-	-	-	Y	-	Y	-	-	0.91800	0.44125	0.60081	0.71864
21	Y	-	-	-	Y	-	-	Y	-	0.88640	0.24957	0.38092	0.62023
22	Y	-	-	-	Y	-	-	-	Y	0.91920	0.44753	0.60651	0.72189
23	Y	-	-	-	-	Y	Y	-	-	0.91680	0.48054	0.61868	0.73443
24	Y	-	-	-	-	Y	-	Y	-	0.88260	0.32308	0.43652	0.64871
25	Y	-	-	-	-	Y	-	-	Y	0.91620	0.47843	0.61628	0.73326
26	Y	-	-	-	-	-	Y	Y	-	0.90120	0.34769	0.49717	0.66986
27	Y	-	-	-	-	-	Y	-	Y	0.95940	0.74429	0.83750	0.86944
28	Y	-	-	-	-	-	-	Y	Y	0.90040	0.34423	0.49331	0.66802
29	-	Y	Y	Y	-	-	-	-	-	0.96140	0.77303	0.84777	0.88254
30	-	Y	Y	-	Y	-	-	-	-	0.94780	0.66754	0.78078	0.83061
31	-	Y	Y	-	-	Y	-	-	-	0.94660	0.70299	0.78781	0.84484
32	-	Y	Y	-	-	-	Y	-	-	0.95940	0.75893	0.84011	0.87560
33	-	Y	Y	-	-	-	-	Y	-	0.94760	0.68321	0.78553	0.83717
34	-	Y	Y	-	-	-	-	-	Y	0.96020	0.76954	0.84461	0.88057
35	-	Y	-	Y	Y	-	-	-	-	0.95360	0.71276	0.81043	0.85274
36	-	Y	-	Y	-	Y	-	-	-	0.95420	0.75523	0.82197	0.87095
37	-	Y	-	Y	-	-	Y	-	-	0.96100	0.77132	0.84775	0.88167
38	-	Y	-	Y	-	-	-	Y	-	0.95400	0.72384	0.81358	0.85771
39	-	Y	-	Y	-	-	-	-	Y	0.96180	0.78429	0.85219	0.88758
40	-	Y	-	-	Y	Y	-	-	-	0.92980	0.58088	0.69926	0.78402
41	-	Y	-	-	Y	-	Y	-	-	0.95780	0.73585	0.83045	0.86500
42	-	Y	-	-	Y	-	-	Y	-	0.92120	0.49290	0.63602	0.74214
43	-	Y	-	-	Y	-	-	-	Y	0.95700	0.73856	0.82802	0.86566
44	-	Y	-	-	-	Y	Y	-	-	0.95660	0.74530	0.82780	0.86821
45	-	Y	-	-	-	Y	-	Y	-	0.91220	0.52713	0.62730	0.75121
46	-	Y	-	-	-	Y	-	-	Y	0.95600	0.74640	0.82552	0.86830
47	-	Y	-	-	-	-	Y	Y	-	0.95540	0.74294	0.82392	0.86656
48	-	Y	-	-	-	-	Y	-	Y	0.95860	0.76247	0.83829	0.87655
49	-	Y	-	-	-	-	-	Y	Y	0.95500	0.74539	0.82274	0.86733
50	-	-	Y	Y	Y	-	-	-	-	0.94840	0.68383	0.78647	0.83769
51	-	-	Y	Y	-	Y	-	-	-	0.94940	0.74179	0.80367	0.86246
52	-	-	Y	Y	-	-	Y	-	-	0.96260	0.77328	0.85083	0.88324
53	-	-	Y	Y	-	-	-	Y	-	0.95220	0.71151	0.80559	0.85154
54	-	-	Y	Y	-	-	-	-	Y	0.96220	0.77377	0.85015	0.88326
55	-	-	Y	-	Y	Y	-	-	-	0.92540	0.56593	0.67938	0.77515
56	-	-	Y	-	Y	-	Y	-	-	0.94980	0.68145	0.79145	0.83755
57	-	-	Y	-	Y	-	-	Y	-	0.92180	0.49348	0.63814	0.74277
58	-	-	Y	-	Y	-	-	-	Y	0.95020	0.68306	0.79266	0.83849
59	-	-	Y	-	-	Y	Y	-	-	0.95060	0.72433	0.80528	0.85608
60	-	-	Y	-	-	Y	-	Y	-	0.90320	0.49011	0.58706	0.73046
61	-	-	Y	-	-	Y	-	-	Y	0.95080	0.72703	0.80582	0.85731
62	-	-	Y	-	-	-	Y	Y	-	0.95200	0.69833	0.80347	0.84599
63	-	-	Y	-	-	-	Y	-	Y	0.96260	0.76987	0.85252	0.88200
64	-	-	Y	-	-	-	-	Y	Y	0.95200	0.70238	0.80353	0.84768
65	-	-	-	Y	Y	Y	-	-	-	0.92680	0.57750	0.68896	0.78069
66	-	-	-	Y	Y	-	Y	-	-	0.95560	0.71947	0.81848	0.85668
67	-	-	-	Y	Y	-	-	Y	-	0.92360	0.50757	0.65017	0.74957
68	-	-	-	Y	Y	-	-	-	Y	0.95600	0.72199	0.82110	0.85805
69	-	-	-	Y	-	Y	Y	-	-	0.95660	0.76620	0.83188	0.87689
70	-	-	-	Y	-	Y	-	Y	-	0.90480	0.49614	0.59474	0.73394
71	-	-	-	Y	-	Y	-	-	Y	0.95800	0.77439	0.83746	0.88110
72	-	-	-	Y	-	-	Y	Y	-	0.95600	0.72339	0.82078	0.85864
73	-	-	-	Y	-	-	Y	-	Y	0.96420	0.78085	0.85984	0.88748
74	-	-	-	Y	-	-	-	Y	Y	0.95720	0.73294	0.82618	0.86330
75	-	-	-	-	Y	Y	Y	-	-	0.93420	0.59880	0.71868	0.79402
76	-	-	-	-	Y	Y	-	Y	-	0.89400	0.41596	0.52326	0.69397
77	-	-	-	-	Y	Y	-	-	Y	0.93440	0.60027	0.71966	0.79476
78	-	-	-	-	Y	-	Y	Y	-	0.92720	0.51658	0.66459	0.75548
79	-	-	-	-	Y	-	Y	-	Y	0.96020	0.75124	0.84161	0.87280
80	-	-	-	-	Y	-	-	Y	Y	0.92720	0.51941	0.66591	0.75666
81	-	-	-	-	-	Y	Y	Y	-	0.91340	0.52477	0.63013	0.75095
82	-	-	-	-	-	Y	Y	-	Y	0.96040	0.75227	0.84189	0.87332
83	-	-	-	-	-	Y	-	Y	Y	0.91320	0.52613	0.62997	0.75140
84	-	-	-	-	-	-	Y	Y	Y	0.95980	0.75092	0.83991	0.87240

Table A.10: Scores of ensembles of size three - UCL

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
1	Y	Y	Y	-	-	-	-	-	-	0.90292	0.75601	0.81918	0.85994
2	Y	Y	-	Y	-	-	-	-	-	0.90377	0.76174	0.82170	0.86218
3	Y	Y	-	-	Y	-	-	-	-	0.89640	0.73778	0.80587	0.85000
4	Y	Y	-	-	-	Y	-	-	-	0.84992	0.59806	0.69853	0.77607
5	Y	Y	-	-	-	-	Y	-	-	0.91327	0.77168	0.83832	0.87186
6	Y	Y	-	-	-	-	-	Y	-	0.90462	0.75274	0.82155	0.86028
7	Y	Y	-	-	-	-	-	-	Y	0.92106	0.79979	0.85503	0.88551
8	Y	-	Y	Y	-	-	-	-	-	0.90518	0.79653	0.82964	0.87329
9	Y	-	Y	-	Y	-	-	-	-	0.89909	0.76722	0.81530	0.86048
10	Y	-	Y	-	-	Y	-	-	-	0.85275	0.61428	0.70776	0.78281
11	Y	-	Y	-	-	-	Y	-	-	0.91270	0.77666	0.83785	0.87274
12	Y	-	Y	-	-	-	-	Y	-	0.90561	0.76814	0.82543	0.86528
13	Y	-	Y	-	-	-	-	-	Y	0.92035	0.80508	0.85436	0.88643
14	Y	-	-	Y	Y	-	-	-	-	0.89909	0.77155	0.81626	0.86183
15	Y	-	-	Y	-	Y	-	-	-	0.85417	0.61329	0.70945	0.78352
16	Y	-	-	Y	-	-	Y	-	-	0.91284	0.78244	0.83911	0.87454
17	Y	-	-	Y	-	-	-	Y	-	0.90306	0.76212	0.82059	0.86163
18	Y	-	-	Y	-	-	-	-	Y	0.92149	0.81669	0.85792	0.89064
19	Y	-	-	-	Y	Y	-	-	-	0.84609	0.60301	0.69463	0.77488
20	Y	-	-	-	Y	-	Y	-	-	0.90519	0.76096	0.82363	0.86288
21	Y	-	-	-	Y	-	-	Y	-	0.89909	0.74272	0.81096	0.85323
22	Y	-	-	-	Y	-	-	-	Y	0.91057	0.78328	0.83607	0.87322
23	Y	-	-	-	-	Y	Y	-	-	0.85658	0.61071	0.71207	0.78441
24	Y	-	-	-	-	Y	-	Y	-	0.85275	0.60875	0.70604	0.78114
25	Y	-	-	-	-	Y	-	-	Y	0.85998	0.62138	0.72037	0.78994
26	Y	-	-	-	-	-	Y	Y	-	0.91313	0.77230	0.83832	0.87182
27	Y	-	-	-	-	-	Y	-	Y	0.92971	0.82431	0.87217	0.89876
28	Y	-	-	-	-	-	-	Y	Y	0.92120	0.79800	0.85525	0.88505
29	-	Y	Y	Y	-	-	-	-	-	0.91454	0.83460	0.85006	0.89114
30	-	Y	Y	-	Y	-	-	-	-	0.91298	0.81381	0.84457	0.88393
31	-	Y	Y	-	-	Y	-	-	-	0.90207	0.75120	0.81714	0.85790
32	-	Y	Y	-	-	-	Y	-	-	0.92106	0.81483	0.85730	0.88992
33	-	Y	Y	-	-	-	-	Y	-	0.91780	0.81223	0.85197	0.88693
34	-	Y	Y	-	-	-	-	-	Y	0.93084	0.85072	0.87749	0.90734
35	-	Y	-	Y	Y	-	-	-	-	0.91610	0.82956	0.85175	0.89081
36	-	Y	-	Y	-	Y	-	-	-	0.90320	0.75214	0.81901	0.85897
37	-	Y	-	Y	-	-	Y	-	-	0.92319	0.82551	0.86218	0.89455
38	-	Y	-	Y	-	-	-	Y	-	0.91894	0.81882	0.85473	0.88960
39	-	Y	-	Y	-	-	-	-	Y	0.93112	0.85728	0.87865	0.90941
40	-	Y	-	-	Y	Y	-	-	-	0.88946	0.71638	0.79063	0.83879
41	-	Y	-	-	Y	-	Y	-	-	0.92092	0.80852	0.85627	0.88795
42	-	Y	-	-	Y	-	-	Y	-	0.91582	0.79687	0.84655	0.88091
43	-	Y	-	-	Y	-	-	-	Y	0.92871	0.84117	0.87314	0.90305
44	-	Y	-	-	-	Y	Y	-	-	0.91199	0.76228	0.83472	0.86814
45	-	Y	-	-	-	Y	-	Y	-	0.90377	0.74185	0.81818	0.85628
46	-	Y	-	-	-	Y	-	-	Y	0.91978	0.79062	0.85176	0.88190
47	-	Y	-	-	-	-	Y	Y	-	0.92375	0.80864	0.86061	0.88991
48	-	Y	-	-	-	-	Y	-	Y	0.93268	0.84400	0.87942	0.90659
49	-	Y	-	-	-	-	-	Y	Y	0.92886	0.83376	0.87231	0.90082
50	-	-	Y	Y	Y	-	-	-	-	0.91142	0.84053	0.84621	0.89072
51	-	-	Y	Y	-	Y	-	-	-	0.90547	0.79573	0.83010	0.87329
52	-	-	Y	Y	-	-	Y	-	-	0.92063	0.84797	0.86099	0.89932
53	-	-	Y	Y	-	-	-	Y	-	0.91567	0.83387	0.85179	0.89175
54	-	-	Y	Y	-	-	-	-	Y	0.92474	0.86447	0.86961	0.90704
55	-	-	Y	-	Y	Y	-	-	-	0.89385	0.75112	0.80455	0.85209
56	-	-	Y	-	Y	-	Y	-	-	0.91907	0.82190	0.85484	0.89062
57	-	-	Y	-	Y	-	-	Y	-	0.91440	0.81682	0.84695	0.88575
58	-	-	Y	-	Y	-	-	-	Y	0.92588	0.84744	0.86896	0.90284
59	-	-	Y	-	-	Y	Y	-	-	0.91227	0.77057	0.83630	0.87068
60	-	-	Y	-	-	Y	-	Y	-	0.90561	0.76377	0.82480	0.86395
61	-	-	Y	-	-	Y	-	-	Y	0.92134	0.80472	0.85606	0.88703
62	-	-	Y	-	-	-	Y	Y	-	0.92219	0.81474	0.85909	0.89066
63	-	-	Y	-	-	-	Y	-	Y	0.93537	0.85880	0.88550	0.91288
64	-	-	Y	-	-	-	-	Y	Y	0.92914	0.84590	0.87439	0.90470
65	-	-	-	Y	Y	Y	-	-	-	0.89442	0.75414	0.80599	0.85341
66	-	-	-	Y	Y	-	Y	-	-	0.92290	0.83842	0.86314	0.89818
67	-	-	-	Y	Y	-	-	Y	-	0.91695	0.83064	0.85294	0.89166
68	-	-	-	Y	Y	-	-	-	Y	0.92956	0.86686	0.87697	0.91119
69	-	-	-	Y	-	Y	Y	-	-	0.91270	0.77302	0.83713	0.87165
70	-	-	-	Y	-	Y	-	Y	-	0.90391	0.75553	0.82041	0.86025
71	-	-	-	Y	-	Y	-	-	Y	0.92191	0.81309	0.85793	0.88983
72	-	-	-	Y	-	-	Y	Y	-	0.92460	0.82692	0.86436	0.89587
73	-	-	-	Y	-	-	Y	-	Y	0.93637	0.86713	0.88772	0.91597
74	-	-	-	Y	-	-	-	Y	Y	0.93084	0.85602	0.87794	0.90878
75	-	-	-	-	Y	Y	Y	-	-	0.89895	0.74224	0.81053	0.85301
76	-	-	-	-	Y	Y	-	Y	-	0.89272	0.72462	0.79725	0.84338
77	-	-	-	-	Y	Y	-	-	Y	0.90320	0.76183	0.82107	0.86178
78	-	-	-	-	Y	-	Y	Y	-	0.92205	0.80961	0.85795	0.88901
79	-	-	-	-	Y	-	Y	-	Y	0.93367	0.84964	0.88185	0.90899
80	-	-	-	-	Y	-	-	Y	Y	0.93127	0.84239	0.87710	0.90516
81	-	-	-	-	-	Y	Y	Y	-	0.91298	0.76503	0.83657	0.86949
82	-	-	-	-	-	Y	Y	-	Y	0.92829	0.81452	0.86859	0.89485
83	-	-	-	-	-	Y	-	Y	Y	0.92078	0.79057	0.85321	0.88243
84	-	-	-	-	-	-	Y	Y	Y	0.93297	0.84013	0.87950	0.90563

Table A.11: Scores of ensembles of size three - Chile

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
1	Y	Y	Y	-	-	-	-	-	-	0.77999	0.08119	0.14707	0.53746
2	Y	Y	-	Y	-	-	-	-	-	0.78523	0.13068	0.22090	0.55803
3	Y	Y	-	-	Y	-	-	-	-	0.78711	0.13399	0.22693	0.56038
4	Y	Y	-	-	-	Y	-	-	-	0.78066	0.17761	0.27360	0.57120
5	Y	Y	-	-	-	-	Y	-	-	0.78838	0.14243	0.23876	0.56412
6	Y	Y	-	-	-	-	-	Y	-	0.77541	0.07658	0.13706	0.53284
7	Y	Y	-	-	-	-	-	-	Y	0.78936	0.14825	0.24696	0.56677
8	Y	-	Y	Y	-	-	-	-	-	0.78186	0.09661	0.17160	0.54409
9	Y	-	Y	-	Y	-	-	-	-	0.78246	0.09757	0.17343	0.54481
10	Y	-	Y	-	-	Y	-	-	-	0.77736	0.10829	0.18535	0.54518
11	Y	-	Y	-	-	-	Y	-	-	0.78261	0.09269	0.16623	0.54321
12	Y	-	Y	-	-	-	-	Y	-	0.77136	0.04709	0.08756	0.52001
13	Y	-	Y	-	-	-	-	-	Y	0.78253	0.09286	0.16642	0.54319
14	Y	-	-	Y	Y	-	-	-	-	0.78988	0.19565	0.30336	0.58367
15	Y	-	-	Y	-	Y	-	-	-	0.77526	0.23966	0.33265	0.58936
16	Y	-	-	Y	-	-	Y	-	-	0.79490	0.18440	0.29571	0.58304
17	Y	-	-	Y	-	-	-	Y	-	0.77474	0.07667	0.13692	0.53250
18	Y	-	-	Y	-	-	-	-	Y	0.79393	0.17957	0.28931	0.58073
19	Y	-	-	-	Y	Y	-	-	-	0.75667	0.34767	0.40087	0.61476
20	Y	-	-	-	Y	-	Y	-	-	0.78681	0.25440	0.35835	0.60210
21	Y	-	-	-	Y	-	-	Y	-	0.77451	0.07302	0.13133	0.53107
22	Y	-	-	-	Y	-	-	-	Y	0.79865	0.20281	0.32009	0.59186
23	Y	-	-	-	-	Y	Y	-	-	0.77541	0.31993	0.40011	0.61738
24	Y	-	-	-	-	Y	-	Y	-	0.77219	0.09649	0.16484	0.53767
25	Y	-	-	-	-	Y	-	-	Y	0.79273	0.26162	0.37128	0.60838
26	Y	-	-	-	-	-	Y	Y	-	0.77541	0.07646	0.13693	0.53284
27	Y	-	-	-	-	-	Y	-	Y	0.80022	0.21149	0.33105	0.59591
28	Y	-	-	-	-	-	-	Y	Y	0.77519	0.08137	0.14442	0.53437
29	-	Y	Y	Y	-	-	-	-	-	0.78801	0.14661	0.24399	0.56540
30	-	Y	Y	-	Y	-	-	-	-	0.79018	0.15282	0.25386	0.56897
31	-	Y	Y	-	-	Y	-	-	-	0.78643	0.19701	0.30072	0.58173
32	-	Y	Y	-	-	-	Y	-	-	0.79160	0.15385	0.25607	0.57022
33	-	Y	Y	-	-	-	-	Y	-	0.78553	0.10835	0.19091	0.55050
34	-	Y	Y	-	-	-	-	-	Y	0.79265	0.16121	0.26621	0.57345
35	-	Y	-	Y	Y	-	-	-	-	0.79220	0.22258	0.33357	0.59449
36	-	Y	-	Y	-	Y	-	-	-	0.77984	0.27842	0.37138	0.60569
37	-	Y	-	Y	-	-	Y	-	-	0.79655	0.21216	0.32742	0.59368
38	-	Y	-	Y	-	-	-	Y	-	0.78861	0.14513	0.24260	0.56525
39	-	Y	-	Y	-	-	-	-	Y	0.79700	0.21055	0.32625	0.59342
40	-	Y	-	-	Y	Y	-	-	-	0.75952	0.38050	0.42563	0.62792
41	-	Y	-	-	Y	-	Y	-	-	0.78883	0.27870	0.38194	0.61180
42	-	Y	-	-	Y	-	-	Y	-	0.78921	0.14820	0.24711	0.56670
43	-	Y	-	-	Y	-	-	-	Y	0.80120	0.23164	0.35257	0.60348
44	-	Y	-	-	-	Y	Y	-	-	0.77699	0.34568	0.42057	0.62730
45	-	Y	-	-	-	Y	-	Y	-	0.78253	0.18366	0.28271	0.57458
46	-	Y	-	-	-	Y	-	-	Y	0.79363	0.28140	0.38953	0.61576
47	-	Y	-	-	-	-	Y	Y	-	0.78988	0.15191	0.25205	0.56841
48	-	Y	-	-	-	-	Y	-	Y	0.80105	0.23234	0.35307	0.60363
49	-	Y	-	-	-	-	-	Y	Y	0.78988	0.15455	0.25559	0.56934
50	-	-	Y	Y	Y	-	-	-	-	0.79063	0.19969	0.30845	0.58555
51	-	-	Y	Y	-	Y	-	-	-	0.77961	0.23944	0.33669	0.59214
52	-	-	Y	Y	-	-	Y	-	-	0.79550	0.18952	0.30220	0.58521
53	-	-	Y	Y	-	-	-	Y	-	0.78628	0.12540	0.21513	0.55697
54	-	-	Y	Y	-	-	-	-	Y	0.79460	0.18570	0.29698	0.58331
55	-	-	Y	-	Y	Y	-	-	-	0.76012	0.34636	0.40339	0.61656
56	-	-	Y	-	Y	-	Y	-	-	0.78726	0.26049	0.36431	0.60450
57	-	-	Y	-	Y	-	-	Y	-	0.78696	0.12224	0.21164	0.55632
58	-	-	Y	-	Y	-	-	-	Y	0.79948	0.21059	0.32920	0.59511
59	-	-	Y	-	-	Y	Y	-	-	0.77871	0.32489	0.40729	0.62129
60	-	-	Y	-	-	Y	-	Y	-	0.78253	0.14503	0.23753	0.56130
61	-	-	Y	-	-	Y	-	-	Y	0.79670	0.26811	0.38166	0.61325
62	-	-	Y	-	-	-	Y	Y	-	0.78726	0.11880	0.20689	0.55529
63	-	-	Y	-	-	-	Y	-	Y	0.80105	0.21577	0.33637	0.59795
64	-	-	Y	-	-	-	-	Y	Y	0.78696	0.12409	0.21394	0.55691
65	-	-	-	Y	Y	Y	-	-	-	0.75105	0.38046	0.41708	0.62248
66	-	-	-	Y	Y	-	Y	-	-	0.78403	0.28944	0.38563	0.61246
67	-	-	-	Y	Y	-	-	Y	-	0.79130	0.20478	0.31437	0.58775
68	-	-	-	Y	Y	-	-	-	Y	0.79595	0.24705	0.36135	0.60547
69	-	-	-	Y	-	Y	Y	-	-	0.76462	0.36527	0.42082	0.62609
70	-	-	-	Y	-	Y	-	Y	-	0.77781	0.24636	0.34125	0.59334
71	-	-	-	Y	-	Y	-	-	Y	0.78396	0.31418	0.40512	0.62088
72	-	-	-	Y	-	-	Y	Y	-	0.79678	0.19517	0.30991	0.58799
73	-	-	-	Y	-	-	Y	-	Y	0.80105	0.24519	0.36578	0.60816
74	-	-	-	Y	-	-	-	Y	Y	0.79603	0.19098	0.30417	0.58604
75	-	-	-	-	Y	Y	Y	-	-	0.75457	0.40459	0.43565	0.63320
76	-	-	-	-	Y	Y	-	Y	-	0.75915	0.35727	0.40984	0.61971
77	-	-	-	-	Y	Y	-	-	Y	0.76049	0.39757	0.43754	0.63457
78	-	-	-	-	Y	-	Y	Y	-	0.78756	0.26244	0.36646	0.60538
79	-	-	-	-	Y	-	Y	-	Y	0.78868	0.28936	0.39073	0.61546
80	-	-	-	-	Y	-	-	Y	Y	0.79978	0.21534	0.33436	0.59696
81	-	-	-	-	-	Y	Y	Y	-	0.77759	0.32812	0.40844	0.62168
82	-	-	-	-	-	Y	Y	-	Y	0.77781	0.36132	0.43246	0.63333
83	-	-	-	-	-	Y	-	Y	Y	0.79543	0.26469	0.37711	0.61124
84	-	-	-	-	-	-	Y	Y	Y	0.80135	0.22240	0.34336	0.60043

Table A.12: Scores of ensembles of size three - Korea

A.4 Scores of Ensembles of Size Five

Begin of Table													
	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
1	Y	Y	Y	Y	Y	-	-	-	-	0.96048	0.90681	0.94743	0.95102
2	Y	Y	Y	Y	-	Y	-	-	-	0.95984	0.90132	0.94633	0.94953
3	Y	Y	Y	Y	-	-	Y	-	-	0.96824	0.92420	0.95815	0.96051
4	Y	Y	Y	Y	-	-	-	Y	-	0.96864	0.92339	0.95861	0.96070
5	Y	Y	Y	Y	-	-	-	-	Y	0.96864	0.92423	0.95866	0.96086
6	Y	Y	Y	-	Y	Y	-	-	-	0.95712	0.89763	0.94264	0.94664
7	Y	Y	Y	-	Y	-	Y	-	-	0.96040	0.90700	0.94732	0.95099
8	Y	Y	Y	-	Y	-	-	Y	-	0.96088	0.90702	0.94793	0.95140
9	Y	Y	Y	-	Y	-	-	-	Y	0.96088	0.90764	0.94798	0.95152
10	Y	Y	Y	-	-	Y	Y	-	-	0.95968	0.90151	0.94612	0.94943
11	Y	Y	Y	-	-	Y	-	Y	-	0.96008	0.90070	0.94658	0.94962
12	Y	Y	Y	-	-	Y	-	-	Y	0.95992	0.90111	0.94639	0.94957
13	Y	Y	Y	-	-	-	Y	Y	-	0.96864	0.92460	0.95867	0.96091
14	Y	Y	Y	-	-	-	Y	-	Y	0.96856	0.92523	0.95859	0.96096
15	Y	Y	Y	-	-	-	-	Y	Y	0.96888	0.92420	0.95895	0.96104
16	Y	Y	-	Y	Y	Y	-	-	-	0.96056	0.90457	0.94741	0.95069
17	Y	Y	-	Y	Y	-	Y	-	-	0.96816	0.92543	0.95810	0.96067
18	Y	Y	-	Y	Y	-	-	Y	-	0.96880	0.92444	0.95888	0.96103
19	Y	Y	-	Y	Y	-	-	-	Y	0.96864	0.92506	0.95870	0.96101
20	Y	Y	-	Y	-	Y	Y	-	-	0.96848	0.92360	0.95843	0.96061
21	Y	Y	-	Y	-	Y	-	Y	-	0.96856	0.92339	0.95852	0.96063
22	Y	Y	-	Y	-	Y	-	-	Y	0.96856	0.92423	0.95857	0.96079
23	Y	Y	-	Y	-	-	Y	Y	-	0.96920	0.92524	0.95942	0.96149
24	Y	Y	-	Y	-	-	Y	-	Y	0.96896	0.92607	0.95915	0.96145
25	Y	Y	-	Y	-	-	-	Y	Y	0.96952	0.92607	0.95986	0.96191
26	Y	Y	-	-	Y	Y	Y	-	-	0.96048	0.90457	0.94730	0.95063
27	Y	Y	-	-	Y	Y	-	Y	-	0.96104	0.90436	0.94800	0.95105
28	Y	Y	-	-	Y	Y	-	-	Y	0.96080	0.90498	0.94773	0.95097
29	Y	Y	-	-	Y	-	Y	Y	-	0.96856	0.92504	0.95860	0.96093
30	Y	Y	-	-	Y	-	Y	-	Y	0.96840	0.92629	0.95846	0.96103
31	Y	Y	-	-	Y	-	-	Y	Y	0.96896	0.92525	0.95911	0.96130
32	Y	Y	-	-	-	Y	Y	Y	-	0.96880	0.92461	0.95888	0.96105
33	Y	Y	-	-	-	Y	Y	-	Y	0.96880	0.92504	0.95890	0.96113
34	Y	Y	-	-	-	Y	-	Y	Y	0.96904	0.92482	0.95918	0.96129
35	Y	Y	-	-	-	-	Y	Y	Y	0.96920	0.92522	0.95940	0.96149
36	Y	-	Y	Y	Y	Y	-	-	-	0.95672	0.89784	0.94218	0.94634
37	Y	-	Y	Y	Y	-	Y	-	-	0.96008	0.90779	0.94696	0.95086
38	Y	-	Y	Y	Y	-	-	Y	-	0.96064	0.90722	0.94765	0.95123
39	Y	-	Y	Y	Y	-	-	-	Y	0.96064	0.90763	0.94767	0.95131
40	Y	-	Y	Y	-	Y	Y	-	-	0.95952	0.90192	0.94595	0.94936
41	Y	-	Y	Y	-	Y	-	Y	-	0.95992	0.90153	0.94644	0.94964
42	Y	-	Y	Y	-	Y	-	-	Y	0.95992	0.90195	0.94647	0.94971
43	Y	-	Y	Y	-	-	Y	Y	-	0.96840	0.92459	0.95835	0.96070
44	Y	-	Y	Y	-	-	Y	-	Y	0.96832	0.92542	0.95829	0.96079
45	Y	-	Y	Y	-	-	-	Y	Y	0.96880	0.92482	0.95888	0.96109
46	Y	-	Y	-	Y	Y	Y	-	-	0.95640	0.89719	0.94169	0.94596
47	Y	-	Y	-	Y	Y	-	Y	-	0.95720	0.89763	0.94274	0.94671
48	Y	-	Y	-	Y	Y	-	-	Y	0.95696	0.89805	0.94246	0.94659
49	Y	-	Y	-	Y	-	Y	Y	-	0.96056	0.90719	0.94752	0.95116
50	Y	-	Y	-	Y	-	Y	-	Y	0.96032	0.90803	0.94727	0.95112
51	Y	-	Y	-	Y	-	-	Y	Y	0.96096	0.90784	0.94809	0.95161
52	Y	-	Y	-	-	Y	Y	Y	-	0.95984	0.90169	0.94632	0.94959
53	Y	-	Y	-	-	Y	Y	-	Y	0.95960	0.90211	0.94603	0.94948
54	Y	-	Y	-	-	Y	-	Y	Y	0.96008	0.90150	0.94661	0.94976
55	Y	-	Y	-	-	-	Y	Y	Y	0.96880	0.92582	0.95891	0.96126
56	Y	-	-	Y	Y	Y	Y	-	-	0.95992	0.90576	0.94664	0.95037
57	Y	-	-	Y	Y	Y	-	Y	-	0.96088	0.90519	0.94785	0.95107
58	Y	-	-	Y	Y	Y	-	-	Y	0.96048	0.90560	0.94736	0.95082
59	Y	-	-	Y	Y	-	Y	Y	-	0.96848	0.92603	0.95853	0.96103
60	Y	-	-	Y	Y	-	Y	-	Y	0.96824	0.92666	0.95824	0.96096
61	Y	-	-	Y	Y	-	-	Y	Y	0.96912	0.92586	0.95933	0.96155
62	Y	-	-	Y	-	Y	Y	Y	-	0.96896	0.92460	0.95907	0.96117
63	Y	-	-	Y	-	Y	Y	-	Y	0.96880	0.92586	0.95893	0.96128
64	Y	-	-	Y	-	Y	-	Y	Y	0.96904	0.92502	0.95920	0.96132
65	Y	-	-	Y	-	-	Y	Y	Y	0.96920	0.92626	0.95946	0.96168
66	Y	-	-	-	Y	Y	Y	Y	-	0.96072	0.90517	0.94763	0.95093
67	Y	-	-	-	Y	Y	Y	-	Y	0.96064	0.90642	0.94761	0.95110
68	Y	-	-	-	Y	Y	-	Y	Y	0.96120	0.90579	0.94826	0.95145
69	Y	-	-	-	Y	-	Y	Y	Y	0.96856	0.92647	0.95865	0.96119
70	Y	-	-	-	-	Y	Y	Y	Y	0.96904	0.92543	0.95920	0.96140
71	-	Y	Y	Y	Y	Y	-	-	-	0.96168	0.90701	0.94895	0.95205
72	-	Y	Y	Y	Y	-	Y	-	-	0.96808	0.92481	0.95795	0.96049
73	-	Y	Y	Y	Y	-	-	Y	-	0.96864	0.92360	0.95862	0.96074
74	-	Y	Y	Y	Y	-	-	-	Y	0.96856	0.92485	0.95859	0.96090
75	-	Y	Y	Y	-	Y	Y	-	-	0.96848	0.92339	0.95841	0.96057
76	-	Y	Y	Y	-	Y	-	Y	-	0.96864	0.92319	0.95860	0.96066
77	-	Y	Y	Y	-	Y	-	-	Y	0.96856	0.92381	0.95853	0.96071
78	-	Y	Y	Y	-	-	Y	Y	-	0.96896	0.92441	0.95907	0.96114

Continuation of Table

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
79	-	Y	Y	Y	-	-	Y	-	Y	0.96864	0.92503	0.95869	0.96099
80	-	Y	Y	Y	-	-	-	Y	Y	0.96928	0.92504	0.95950	0.96153
81	-	Y	Y	-	Y	Y	Y	-	-	0.96160	0.90722	0.94886	0.95202
82	-	Y	Y	-	Y	Y	-	Y	-	0.96224	0.90723	0.94968	0.95255
83	-	Y	Y	-	Y	Y	-	-	Y	0.96192	0.90764	0.94929	0.95237
84	-	Y	Y	-	Y	-	Y	Y	-	0.96832	0.92400	0.95823	0.96054
85	-	Y	Y	-	Y	-	Y	-	Y	0.96808	0.92504	0.95797	0.96054
86	-	Y	Y	-	Y	-	-	Y	Y	0.96872	0.92421	0.95874	0.96092
87	-	Y	Y	-	-	Y	Y	Y	-	0.96864	0.92400	0.95864	0.96080
88	-	Y	Y	-	-	Y	Y	-	Y	0.96864	0.92463	0.95867	0.96093
89	-	Y	Y	-	-	Y	-	Y	Y	0.96888	0.92420	0.95895	0.96104
90	-	Y	Y	-	-	-	Y	Y	Y	0.96920	0.92502	0.95939	0.96145
91	-	Y	-	Y	Y	Y	Y	-	-	0.96784	0.92418	0.95762	0.96018
92	-	Y	-	Y	Y	Y	-	Y	-	0.96848	0.92381	0.95843	0.96065
93	-	Y	-	Y	Y	Y	-	-	Y	0.96832	0.92506	0.95830	0.96075
94	-	Y	-	Y	Y	-	Y	Y	-	0.96864	0.92521	0.95868	0.96102
95	-	Y	-	Y	Y	-	Y	-	Y	0.96848	0.92625	0.95853	0.96108
96	-	Y	-	Y	Y	-	-	Y	Y	0.96920	0.92608	0.95946	0.96165
97	-	Y	-	Y	-	Y	Y	Y	-	0.96928	0.92565	0.95953	0.96163
98	-	Y	-	Y	-	Y	Y	-	Y	0.96896	0.92607	0.95914	0.96145
99	-	Y	-	Y	-	Y	-	Y	Y	0.96928	0.92545	0.95952	0.96160
100	-	Y	-	Y	-	-	Y	Y	Y	0.96912	0.92564	0.95932	0.96150
101	-	Y	-	-	Y	Y	Y	Y	-	0.96848	0.92441	0.95846	0.96075
102	-	Y	-	-	Y	Y	Y	-	Y	0.96848	0.92526	0.95850	0.96091
103	-	Y	-	-	Y	Y	-	Y	Y	0.96872	0.92483	0.95878	0.96103
104	-	Y	-	-	Y	-	Y	Y	Y	0.96888	0.92564	0.95902	0.96130
105	-	Y	-	-	-	Y	Y	Y	Y	0.96912	0.92564	0.95932	0.96150
106	-	-	Y	Y	Y	Y	Y	-	-	0.96112	0.90821	0.94827	0.95180
107	-	-	Y	Y	Y	Y	-	Y	-	0.96200	0.90763	0.94939	0.95242
108	-	-	Y	Y	Y	Y	-	-	Y	0.96168	0.90846	0.94903	0.95232
109	-	-	Y	Y	Y	-	Y	Y	-	0.96824	0.92500	0.95816	0.96065
110	-	-	Y	Y	Y	-	Y	-	Y	0.96808	0.92604	0.95801	0.96072
111	-	-	Y	Y	Y	-	-	Y	Y	0.96888	0.92524	0.95900	0.96123
112	-	-	Y	Y	-	Y	Y	Y	-	0.96880	0.92398	0.95883	0.96093
113	-	-	Y	Y	-	Y	Y	-	Y	0.96856	0.92482	0.95856	0.96089
114	-	-	Y	Y	-	Y	-	Y	Y	0.96888	0.92420	0.95894	0.96104
115	-	-	Y	Y	-	-	Y	Y	Y	0.96904	0.92563	0.95921	0.96143
116	-	-	Y	-	Y	Y	Y	Y	-	0.96168	0.90741	0.94896	0.95212
117	-	-	Y	-	Y	Y	Y	-	Y	0.96152	0.90825	0.94881	0.95215
118	-	-	Y	-	Y	Y	-	Y	Y	0.96216	0.90804	0.94960	0.95264
119	-	-	Y	-	Y	-	Y	Y	Y	0.96840	0.92563	0.95839	0.96091
120	-	-	Y	-	-	Y	Y	Y	Y	0.96904	0.92543	0.95920	0.96140
121	-	-	-	Y	Y	Y	Y	Y	-	0.96816	0.92499	0.95805	0.96058
122	-	-	-	Y	Y	Y	Y	-	Y	0.96832	0.92646	0.95834	0.96099
123	-	-	-	Y	Y	Y	-	Y	Y	0.96856	0.92566	0.95862	0.96104
124	-	-	-	Y	Y	-	Y	Y	Y	0.96872	0.92685	0.95886	0.96138
125	-	-	-	Y	-	Y	Y	Y	Y	0.96896	0.92605	0.95913	0.96144
126	-	-	-	-	Y	Y	Y	Y	Y	0.96864	0.92564	0.95871	0.96111

End of Table

Table A.13: Scores of ensembles of size five - Duke

Begin of Table

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
1	Y	Y	Y	Y	Y	-	-	-	-	0.94660	0.64996	0.77150	0.82253
2	Y	Y	Y	Y	-	Y	-	-	-	0.94820	0.68458	0.78707	0.83796
3	Y	Y	Y	Y	-	-	Y	-	-	0.95960	0.74030	0.83574	0.86781
4	Y	Y	Y	Y	-	-	-	Y	-	0.94820	0.66422	0.78199	0.82954
5	Y	Y	Y	Y	-	-	-	-	Y	0.96040	0.74989	0.84086	0.87236
6	Y	Y	Y	-	Y	Y	-	-	-	0.92900	0.54079	0.68092	0.76678
7	Y	Y	Y	-	Y	-	Y	-	-	0.94960	0.66778	0.78744	0.83178
8	Y	Y	Y	-	Y	-	-	Y	-	0.92240	0.47966	0.63439	0.73737
9	Y	Y	Y	-	Y	-	-	-	Y	0.94960	0.66603	0.78619	0.83103
10	Y	Y	Y	-	-	Y	Y	-	-	0.95260	0.69878	0.80572	0.84659
11	Y	Y	Y	-	-	Y	-	Y	-	0.91720	0.47738	0.61691	0.73332
12	Y	Y	Y	-	-	Y	-	-	Y	0.95200	0.69549	0.80239	0.84483
13	Y	Y	Y	-	-	-	Y	Y	-	0.95020	0.67743	0.79292	0.83625
14	Y	Y	Y	-	-	-	Y	-	Y	0.96080	0.75602	0.84465	0.87519
15	Y	Y	Y	-	-	-	-	Y	Y	0.94940	0.67549	0.78903	0.83494
16	Y	Y	-	Y	Y	Y	-	-	-	0.93060	0.55511	0.69181	0.77358
17	Y	Y	-	Y	Y	-	Y	-	-	0.95520	0.70545	0.81394	0.85061
18	Y	Y	-	Y	Y	-	-	Y	-	0.92440	0.48865	0.64462	0.74221
19	Y	Y	-	Y	Y	-	-	-	Y	0.95560	0.70713	0.81585	0.85156
20	Y	Y	-	Y	-	Y	Y	-	-	0.95780	0.73681	0.82991	0.86536
21	Y	Y	-	Y	-	Y	-	Y	-	0.91960	0.49038	0.63105	0.74017
22	Y	Y	-	Y	-	Y	-	-	Y	0.95820	0.73775	0.83098	0.86595
23	Y	Y	-	Y	-	-	Y	Y	-	0.95400	0.70256	0.80996	0.84881
24	Y	Y	-	Y	-	-	Y	-	Y	0.96160	0.76295	0.84883	0.87853
25	Y	Y	-	Y	-	-	-	Y	Y	0.95420	0.70486	0.81075	0.84985
26	Y	Y	-	-	Y	Y	Y	-	-	0.93620	0.58128	0.71934	0.78795
27	Y	Y	-	-	Y	Y	-	Y	-	0.90840	0.41509	0.55837	0.70206
28	Y	Y	-	-	Y	Y	-	-	Y	0.93560	0.57696	0.71561	0.78580
29	Y	Y	-	-	Y	-	Y	Y	-	0.92700	0.50868	0.66096	0.75211
30	Y	Y	-	-	Y	-	Y	-	Y	0.95840	0.73310	0.83183	0.86421
31	Y	Y	-	-	Y	-	-	Y	Y	0.92620	0.50571	0.65713	0.75039
32	Y	Y	-	-	-	Y	Y	Y	-	0.92340	0.51332	0.65261	0.75198
33	Y	Y	-	-	-	Y	Y	-	Y	0.95780	0.73273	0.82946	0.86366
34	Y	Y	-	-	-	Y	-	Y	Y	0.92320	0.51506	0.65289	0.75263
35	Y	Y	-	-	-	-	Y	Y	Y	0.95720	0.73432	0.82821	0.86399
36	Y	-	Y	Y	Y	Y	-	-	-	0.92860	0.55128	0.68396	0.77084
37	Y	-	Y	Y	Y	-	Y	-	-	0.94880	0.66403	0.78296	0.82967
38	Y	-	Y	Y	Y	-	-	Y	-	0.92500	0.50050	0.65155	0.74756
39	Y	-	Y	Y	Y	-	-	-	Y	0.94880	0.66580	0.78342	0.83044
40	Y	-	Y	Y	-	Y	Y	-	-	0.95080	0.69931	0.79916	0.84567
41	Y	-	Y	Y	-	Y	-	Y	-	0.91400	0.46337	0.60192	0.72560
42	Y	-	Y	Y	-	Y	-	-	Y	0.95160	0.70304	0.80186	0.84765
43	Y	-	Y	Y	-	-	Y	Y	-	0.95020	0.67432	0.79136	0.83493
44	Y	-	Y	Y	-	-	Y	-	Y	0.96040	0.74739	0.83973	0.87123
45	Y	-	Y	Y	-	-	-	Y	Y	0.95100	0.67940	0.79437	0.83748
46	Y	-	Y	-	Y	Y	Y	-	-	0.93280	0.56028	0.70019	0.77710
47	Y	-	Y	-	Y	Y	-	Y	-	0.90520	0.39655	0.53869	0.69244
48	Y	-	Y	-	Y	Y	-	-	Y	0.93320	0.56581	0.70297	0.77963
49	Y	-	Y	-	Y	-	Y	Y	-	0.92580	0.49914	0.65358	0.74746
50	Y	-	Y	-	Y	-	Y	-	Y	0.95140	0.67868	0.79603	0.83734
51	Y	-	Y	-	Y	-	-	Y	Y	0.92680	0.50603	0.65930	0.75090
52	Y	-	Y	-	-	Y	Y	Y	-	0.91820	0.47909	0.62140	0.73463
53	Y	-	Y	-	-	Y	Y	-	Y	0.95440	0.71029	0.81427	0.85244
54	Y	-	Y	-	-	Y	-	Y	Y	0.91760	0.47659	0.61857	0.73327
55	Y	-	Y	-	-	-	Y	Y	Y	0.95140	0.68417	0.79802	0.83974
56	Y	-	-	Y	Y	Y	Y	-	-	0.93300	0.56544	0.70323	0.77932
57	Y	-	-	Y	Y	Y	-	Y	-	0.90520	0.39839	0.54072	0.69324
58	Y	-	-	Y	Y	Y	-	-	Y	0.93320	0.56794	0.70420	0.78046
59	Y	-	-	Y	Y	-	Y	Y	-	0.92620	0.50157	0.65593	0.74867
60	Y	-	-	Y	Y	-	Y	-	Y	0.95640	0.71125	0.81937	0.85374
61	Y	-	-	Y	Y	-	-	Y	Y	0.92700	0.50543	0.65981	0.75072
62	Y	-	-	Y	-	Y	Y	Y	-	0.92040	0.49069	0.63392	0.74078
63	Y	-	-	Y	-	Y	Y	-	Y	0.95920	0.74524	0.83630	0.86968
64	Y	-	-	Y	-	Y	-	Y	Y	0.92020	0.49085	0.63353	0.74075
65	Y	-	-	Y	-	-	Y	Y	Y	0.95520	0.70883	0.81501	0.85206
66	Y	-	-	-	Y	Y	Y	Y	-	0.91000	0.41897	0.56469	0.70458
67	Y	-	-	-	Y	Y	Y	-	Y	0.93840	0.59203	0.72946	0.79368
68	Y	-	-	-	Y	Y	-	Y	Y	0.90980	0.41687	0.56352	0.70365
69	Y	-	-	-	Y	-	Y	Y	Y	0.92920	0.52260	0.67351	0.75918
70	Y	-	-	-	-	Y	Y	Y	Y	0.92480	0.51655	0.65872	0.75418
71	-	Y	Y	Y	Y	Y	-	-	-	0.95100	0.70396	0.80067	0.84765
72	-	Y	Y	Y	Y	-	Y	-	-	0.95880	0.74291	0.83373	0.86841
73	-	Y	Y	Y	Y	-	-	Y	-	0.95000	0.67947	0.79143	0.83692
74	-	Y	Y	Y	Y	-	-	-	Y	0.96000	0.75094	0.83996	0.87254
75	-	Y	Y	Y	-	Y	Y	-	-	0.96100	0.76924	0.84644	0.88076
76	-	Y	Y	Y	-	Y	-	Y	-	0.95080	0.71206	0.80214	0.85101
77	-	Y	Y	Y	-	Y	-	-	Y	0.96200	0.77990	0.85198	0.88585
78	-	Y	Y	Y	-	-	Y	Y	-	0.96060	0.75167	0.84136	0.87313
79	-	Y	Y	Y	-	-	Y	-	Y	0.96380	0.78023	0.85832	0.88706
80	-	Y	Y	Y	-	-	-	Y	Y	0.96140	0.76233	0.84627	0.87811
81	-	Y	Y	-	Y	Y	Y	-	-	0.95360	0.71121	0.81189	0.85233
82	-	Y	Y	-	Y	Y	-	Y	-	0.93240	0.57684	0.70521	0.78387

Continuation of Table													
	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
83	-	Y	Y	-	Y	Y	-	-	Y	0.95380	0.70957	0.81150	0.85175
84	-	Y	Y	-	Y	-	Y	Y	-	0.95240	0.69479	0.80356	0.84470
85	-	Y	Y	-	Y	-	Y	-	Y	0.95960	0.75144	0.83959	0.87255
86	-	Y	Y	-	Y	-	-	Y	Y	0.95180	0.69316	0.80068	0.84366
87	-	Y	Y	-	-	Y	Y	Y	-	0.95380	0.71740	0.81416	0.85507
88	-	Y	Y	-	-	Y	Y	-	Y	0.96080	0.76162	0.84546	0.87753
89	-	Y	Y	-	-	Y	-	Y	Y	0.95300	0.71400	0.81019	0.85314
90	-	Y	Y	-	-	-	Y	Y	Y	0.95980	0.75868	0.84176	0.87571
91	-	Y	-	Y	Y	Y	Y	-	-	0.95840	0.74897	0.83428	0.87074
92	-	Y	-	Y	Y	Y	-	Y	-	0.93360	0.58769	0.71351	0.78906
93	-	Y	-	Y	Y	Y	-	-	Y	0.95980	0.75609	0.84018	0.87453
94	-	Y	-	Y	Y	-	Y	Y	-	0.95620	0.72030	0.82086	0.85745
95	-	Y	-	Y	Y	-	Y	-	Y	0.96200	0.76422	0.85015	0.87928
96	-	Y	-	Y	Y	-	-	Y	Y	0.95680	0.72742	0.82450	0.86078
97	-	Y	-	Y	-	Y	Y	Y	-	0.95880	0.75204	0.83597	0.87227
98	-	Y	-	Y	-	Y	Y	-	Y	0.96180	0.76968	0.85042	0.88142
99	-	Y	-	Y	-	Y	-	Y	Y	0.95860	0.75020	0.83450	0.87135
100	-	Y	-	Y	-	-	Y	Y	Y	0.96040	0.76414	0.84503	0.87831
101	-	Y	-	-	Y	Y	Y	Y	-	0.93720	0.60209	0.72903	0.79720
102	-	Y	-	-	Y	Y	Y	-	Y	0.95920	0.74241	0.83612	0.86851
103	-	Y	-	-	Y	Y	-	Y	Y	0.93660	0.60202	0.72699	0.79681
104	-	Y	-	-	Y	-	Y	Y	Y	0.95940	0.74547	0.83767	0.86992
105	-	Y	-	-	-	Y	Y	Y	Y	0.95660	0.73974	0.82675	0.86589
106	-	-	Y	Y	Y	Y	Y	-	-	0.95260	0.71371	0.80804	0.85263
107	-	-	Y	Y	Y	Y	-	Y	-	0.93120	0.57937	0.70253	0.78419
108	-	-	Y	Y	Y	Y	-	-	Y	0.95240	0.71243	0.80695	0.85199
109	-	-	Y	Y	Y	-	Y	Y	-	0.95140	0.68582	0.79763	0.84032
110	-	-	Y	Y	Y	-	Y	-	Y	0.96060	0.74986	0.84110	0.87235
111	-	-	Y	Y	Y	-	-	Y	Y	0.95120	0.68453	0.79656	0.83969
112	-	-	Y	Y	-	Y	Y	Y	-	0.95280	0.72096	0.81085	0.85591
113	-	-	Y	Y	-	Y	Y	-	Y	0.96200	0.77497	0.85073	0.88374
114	-	-	Y	Y	-	Y	-	Y	Y	0.95380	0.72616	0.81455	0.85863
115	-	-	Y	Y	-	-	Y	Y	Y	0.96120	0.75706	0.84413	0.87571
116	-	-	Y	-	Y	Y	Y	Y	-	0.93520	0.59076	0.71861	0.79129
117	-	-	Y	-	Y	Y	Y	-	Y	0.95440	0.71541	0.81567	0.85454
118	-	-	Y	-	Y	Y	-	Y	Y	0.93520	0.59204	0.71886	0.79182
119	-	-	Y	-	Y	-	Y	Y	Y	0.95360	0.69888	0.80882	0.84709
120	-	-	Y	-	-	Y	Y	Y	Y	0.95560	0.72256	0.82062	0.85824
121	-	-	-	Y	Y	Y	Y	Y	-	0.93540	0.59483	0.72158	0.79308
122	-	-	-	Y	Y	Y	Y	-	Y	0.96000	0.75318	0.84058	0.87342
123	-	-	-	Y	Y	Y	-	Y	Y	0.93600	0.59841	0.72378	0.79487
124	-	-	-	Y	Y	-	Y	Y	Y	0.95800	0.72699	0.82831	0.86126
125	-	-	-	Y	-	Y	Y	Y	Y	0.95960	0.75027	0.83810	0.87197
126	-	-	-	-	Y	Y	Y	Y	Y	0.94080	0.61470	0.74434	0.80454
End of Table													

Table A.14: Scores of ensembles of size five - UCL

Begin of Table

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
1	Y	Y	Y	Y	Y	-	-	-	-	0.91270	0.79875	0.84154	0.87932
2	Y	Y	Y	Y	-	Y	-	-	-	0.90377	0.74928	0.81919	0.85856
3	Y	Y	Y	Y	-	-	Y	-	-	0.91539	0.79482	0.84524	0.88001
4	Y	Y	Y	Y	-	-	-	Y	-	0.91355	0.79378	0.84249	0.87848
5	Y	Y	Y	Y	-	-	-	-	Y	0.92389	0.82084	0.86238	0.89362
6	Y	Y	Y	-	Y	Y	-	-	-	0.89739	0.72866	0.80543	0.84803
7	Y	Y	Y	-	Y	-	Y	-	-	0.91482	0.78187	0.84235	0.87583
8	Y	Y	Y	-	Y	-	-	Y	-	0.91355	0.78290	0.84052	0.87525
9	Y	Y	Y	-	Y	-	-	-	Y	0.92049	0.80378	0.85452	0.88619
10	Y	Y	Y	-	-	Y	Y	-	-	0.90873	0.74987	0.82715	0.86225
11	Y	Y	Y	-	-	Y	-	Y	-	0.90377	0.73915	0.81747	0.85556
12	Y	Y	Y	-	-	Y	-	-	Y	0.91213	0.76339	0.83486	0.86850
13	Y	Y	Y	-	-	-	Y	Y	-	0.91978	0.79289	0.85219	0.88263
14	Y	Y	Y	-	-	-	Y	-	Y	0.92786	0.81845	0.86830	0.89573
15	Y	Y	Y	-	-	-	-	Y	Y	0.92644	0.81512	0.86601	0.89384
16	Y	Y	-	Y	Y	Y	-	-	-	0.89683	0.72468	0.80374	0.84644
17	Y	Y	-	Y	Y	-	Y	-	-	0.91766	0.79341	0.84860	0.88122
18	Y	Y	-	Y	Y	-	-	Y	-	0.91398	0.78627	0.84178	0.87654
19	Y	Y	-	Y	Y	-	-	-	Y	0.92276	0.81457	0.85975	0.89099
20	Y	Y	-	Y	-	Y	Y	-	-	0.90844	0.74849	0.82635	0.86156
21	Y	Y	-	Y	-	Y	-	Y	-	0.90235	0.73483	0.81432	0.85318
22	Y	Y	-	Y	-	Y	-	-	Y	0.91284	0.76706	0.83661	0.87003
23	Y	Y	-	Y	-	-	Y	Y	-	0.92007	0.79641	0.85293	0.88381
24	Y	Y	-	Y	-	-	Y	-	Y	0.92786	0.82271	0.86892	0.89697
25	Y	Y	-	Y	-	-	-	Y	Y	0.92475	0.81686	0.86329	0.89301
26	Y	Y	-	-	Y	Y	Y	-	-	0.90278	0.73141	0.81438	0.85259
27	Y	Y	-	-	Y	Y	-	Y	-	0.89782	0.71920	0.80416	0.84547
28	Y	Y	-	-	Y	Y	-	-	Y	0.90476	0.74091	0.81946	0.85672
29	Y	Y	-	-	Y	-	Y	Y	-	0.91653	0.78064	0.84500	0.87671
30	Y	Y	-	-	Y	-	Y	-	Y	0.92602	0.81105	0.86464	0.89231
31	Y	Y	-	-	Y	-	-	Y	Y	0.92333	0.80200	0.85916	0.88777
32	Y	Y	-	-	-	Y	Y	Y	-	0.90788	0.74576	0.82527	0.86036
33	Y	Y	-	-	-	Y	Y	-	Y	0.91908	0.77731	0.84847	0.87757
34	Y	Y	-	-	-	Y	-	Y	Y	0.91298	0.75793	0.83552	0.86754
35	Y	Y	-	-	-	-	Y	Y	Y	0.92744	0.81219	0.86699	0.89359
36	Y	-	Y	Y	Y	Y	-	-	-	0.89980	0.75104	0.81336	0.85626
37	Y	-	Y	Y	Y	-	Y	-	-	0.91808	0.80906	0.85152	0.88618
38	Y	-	Y	Y	Y	-	-	Y	-	0.91312	0.80123	0.84269	0.88035
39	Y	-	Y	Y	Y	-	-	-	Y	0.92248	0.82821	0.86104	0.89484
40	Y	-	Y	Y	-	Y	Y	-	-	0.90873	0.76097	0.82893	0.86541
41	Y	-	Y	Y	-	Y	-	Y	-	0.90334	0.75208	0.81893	0.85894
42	Y	-	Y	Y	-	Y	-	-	Y	0.91511	0.78439	0.84288	0.87670
43	Y	-	Y	Y	-	-	Y	Y	-	0.91752	0.79867	0.84918	0.88263
44	Y	-	Y	Y	-	-	Y	-	Y	0.92772	0.83141	0.86995	0.89938
45	Y	-	Y	Y	-	-	-	Y	Y	0.92219	0.81859	0.85966	0.89176
46	Y	-	Y	-	Y	Y	Y	-	-	0.90164	0.74032	0.81420	0.85435
47	Y	-	Y	-	Y	Y	-	Y	-	0.89895	0.73455	0.80899	0.85075
48	Y	-	Y	-	Y	Y	-	-	Y	0.90547	0.75479	0.82302	0.86128
49	Y	-	Y	-	Y	-	Y	Y	-	0.91652	0.78789	0.84592	0.87873
50	Y	-	Y	-	Y	-	Y	-	Y	0.92290	0.81176	0.85968	0.89026
51	Y	-	Y	-	Y	-	-	Y	Y	0.92233	0.80674	0.85793	0.88836
52	Y	-	Y	-	-	Y	Y	Y	-	0.90873	0.75258	0.82773	0.86285
53	Y	-	Y	-	-	Y	Y	-	Y	0.91667	0.77600	0.84423	0.87539
54	Y	-	Y	-	-	Y	-	Y	Y	0.91426	0.76778	0.83910	0.87125
55	Y	-	Y	-	-	-	Y	Y	Y	0.92772	0.81444	0.86787	0.89450
56	Y	-	-	Y	Y	Y	Y	-	-	0.90037	0.73792	0.81168	0.85277
57	Y	-	-	Y	Y	Y	-	Y	-	0.89711	0.72569	0.80427	0.84683
58	Y	-	-	Y	Y	Y	-	-	Y	0.90476	0.75247	0.82151	0.86013
59	Y	-	-	Y	Y	-	Y	Y	-	0.91695	0.79369	0.84744	0.88075
60	Y	-	-	Y	Y	-	Y	-	Y	0.92588	0.82368	0.86596	0.89584
61	Y	-	-	Y	Y	-	-	Y	Y	0.92375	0.81528	0.86149	0.89193
62	Y	-	-	Y	-	Y	Y	Y	-	0.90717	0.74607	0.82398	0.85980
63	Y	-	-	Y	-	Y	Y	-	Y	0.91624	0.77775	0.84379	0.87558
64	Y	-	-	Y	-	Y	-	Y	Y	0.91270	0.76362	0.83588	0.86887
65	Y	-	-	Y	-	-	Y	Y	Y	0.92900	0.82431	0.87112	0.89823
66	Y	-	-	-	Y	Y	Y	Y	-	0.90079	0.72854	0.81057	0.85024
67	Y	-	-	-	Y	Y	Y	-	Y	0.90845	0.75532	0.82779	0.86354
68	Y	-	-	-	Y	Y	-	Y	Y	0.90405	0.73977	0.81803	0.85584
69	Y	-	-	-	Y	-	Y	Y	Y	0.92503	0.81007	0.86290	0.89133
70	Y	-	-	-	-	Y	Y	Y	Y	0.91653	0.77097	0.84327	0.87375
71	-	Y	Y	Y	Y	Y	-	-	-	0.91341	0.79346	0.84183	0.87827
72	-	Y	Y	Y	Y	-	Y	-	-	0.92262	0.82773	0.86115	0.89481
73	-	Y	Y	Y	Y	-	-	Y	-	0.92035	0.82432	0.85729	0.89219
74	-	Y	Y	Y	Y	-	-	-	Y	0.92786	0.84920	0.87241	0.90481
75	-	Y	Y	Y	-	Y	Y	-	-	0.91624	0.78950	0.84565	0.87906
76	-	Y	Y	Y	-	Y	-	Y	-	0.91454	0.79081	0.84348	0.87829
77	-	Y	Y	Y	-	Y	-	-	Y	0.92333	0.81551	0.86069	0.89165
78	-	Y	Y	Y	-	-	Y	Y	-	0.92304	0.82127	0.86144	0.89323
79	-	Y	Y	Y	-	-	Y	-	Y	0.93254	0.85278	0.88033	0.90906
80	-	Y	Y	Y	-	-	-	Y	Y	0.92956	0.84523	0.87478	0.90480
81	-	Y	Y	-	Y	Y	Y	-	-	0.91468	0.77615	0.84118	0.87408
82	-	Y	Y	-	Y	Y	-	Y	-	0.91369	0.77763	0.84002	0.87378

Continuation of Table													
	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
83	-	Y	Y	-	Y	Y	-	-	Y	0.92021	0.79811	0.85337	0.88434
84	-	Y	Y	-	Y	-	Y	Y	-	0.92304	0.81028	0.85971	0.88993
85	-	Y	Y	-	Y	-	Y	-	Y	0.93126	0.83932	0.87659	0.90424
86	-	Y	Y	-	Y	-	-	Y	Y	0.92999	0.83452	0.87406	0.90193
87	-	Y	Y	-	-	Y	Y	Y	-	0.91964	0.78738	0.85109	0.88086
88	-	Y	Y	-	-	Y	Y	-	Y	0.92829	0.81371	0.86845	0.89465
89	-	Y	Y	-	-	Y	-	Y	Y	0.92574	0.80958	0.86409	0.89166
90	-	Y	Y	-	-	-	Y	Y	Y	0.93155	0.83595	0.87673	0.90345
91	-	Y	-	Y	Y	Y	Y	-	-	0.91737	0.78610	0.84696	0.87887
92	-	Y	-	Y	Y	Y	-	Y	-	0.91327	0.77602	0.83896	0.87300
93	-	Y	-	Y	Y	Y	-	-	Y	0.92319	0.80873	0.85950	0.88958
94	-	Y	-	Y	Y	-	Y	Y	-	0.92418	0.81746	0.86238	0.89284
95	-	Y	-	Y	Y	-	Y	-	Y	0.93211	0.84679	0.87870	0.90700
96	-	Y	-	Y	Y	-	-	Y	Y	0.93098	0.84306	0.87653	0.90511
97	-	Y	-	Y	-	Y	Y	Y	-	0.91964	0.78840	0.85093	0.88111
98	-	Y	-	Y	-	Y	Y	-	Y	0.92914	0.82034	0.87060	0.89720
99	-	Y	-	Y	-	Y	-	Y	Y	0.92460	0.81047	0.86211	0.89102
100	-	Y	-	Y	-	-	Y	Y	Y	0.93212	0.84106	0.87818	0.90531
101	-	Y	-	-	Y	Y	Y	Y	-	0.91766	0.77614	0.84616	0.87614
102	-	Y	-	-	Y	Y	Y	-	Y	0.92488	0.80384	0.86187	0.88941
103	-	Y	-	-	Y	Y	-	Y	Y	0.92219	0.79305	0.85605	0.88428
104	-	Y	-	-	Y	-	Y	Y	Y	0.93183	0.83112	0.87658	0.90224
105	-	Y	-	-	-	Y	Y	Y	Y	0.92758	0.80805	0.86653	0.89242
106	-	-	Y	Y	Y	Y	Y	-	-	0.91780	0.80378	0.85029	0.88445
107	-	-	Y	Y	Y	Y	-	Y	-	0.91454	0.80007	0.84461	0.88098
108	-	-	Y	Y	Y	Y	-	-	Y	0.92262	0.82586	0.86096	0.89427
109	-	-	Y	Y	Y	-	Y	Y	-	0.92460	0.83488	0.86523	0.89828
110	-	-	Y	Y	Y	-	Y	-	Y	0.93140	0.85770	0.87872	0.90980
111	-	-	Y	Y	Y	-	-	Y	Y	0.92758	0.85002	0.87198	0.90482
112	-	-	Y	Y	-	Y	Y	Y	-	0.91837	0.79660	0.85008	0.88261
113	-	-	Y	Y	-	Y	Y	-	Y	0.92786	0.82737	0.86954	0.89826
114	-	-	Y	Y	-	Y	-	Y	Y	0.92219	0.81501	0.85910	0.89066
115	-	-	Y	Y	-	-	Y	Y	Y	0.93112	0.84843	0.87766	0.90676
116	-	-	Y	-	Y	Y	Y	Y	-	0.91723	0.78535	0.84652	0.87847
117	-	-	Y	-	Y	Y	Y	-	Y	0.92233	0.80652	0.85817	0.88832
118	-	-	Y	-	Y	Y	-	Y	Y	0.92262	0.80421	0.85801	0.88778
119	-	-	Y	-	Y	-	Y	Y	Y	0.93055	0.83424	0.87501	0.90229
120	-	-	Y	-	-	Y	Y	Y	Y	0.92630	0.80747	0.86460	0.89141
121	-	-	-	Y	Y	Y	Y	Y	-	0.91681	0.78680	0.84609	0.87861
122	-	-	-	Y	Y	Y	Y	-	Y	0.92560	0.81718	0.86453	0.89369
123	-	-	-	Y	Y	Y	-	Y	Y	0.92361	0.80814	0.86008	0.88967
124	-	-	-	Y	Y	-	Y	Y	Y	0.93339	0.84693	0.88085	0.90795
125	-	-	-	Y	-	Y	Y	Y	Y	0.92772	0.81673	0.86788	0.89506
126	-	-	-	-	Y	Y	Y	Y	Y	0.92432	0.80215	0.86064	0.88846
End of Table													

Table A.15: Scores of ensembles of size five - Chile

Begin of Table

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
1	Y	Y	Y	Y	Y	-	-	-	-	0.78876	0.13537	0.23047	0.56200
2	Y	Y	Y	Y	-	Y	-	-	-	0.78733	0.15128	0.24938	0.56657
3	Y	Y	Y	Y	-	-	Y	-	-	0.79003	0.13373	0.22920	0.56225
4	Y	Y	Y	Y	-	-	-	Y	-	0.78426	0.09942	0.17714	0.54657
5	Y	Y	Y	Y	-	-	-	-	Y	0.79025	0.13527	0.23138	0.56292
6	Y	Y	Y	-	Y	Y	-	-	-	0.78853	0.15652	0.25684	0.56915
7	Y	Y	Y	-	Y	-	Y	-	-	0.79048	0.13933	0.23681	0.56447
8	Y	Y	Y	-	Y	-	-	Y	-	0.78396	0.09889	0.17629	0.54621
9	Y	Y	Y	-	Y	-	-	-	Y	0.79063	0.14058	0.23873	0.56500
10	Y	Y	Y	-	-	Y	Y	-	-	0.78996	0.15917	0.26113	0.57097
11	Y	Y	Y	-	-	Y	-	Y	-	0.78133	0.11391	0.19542	0.54965
12	Y	Y	Y	-	-	Y	-	-	Y	0.79100	0.16461	0.26868	0.57353
13	Y	Y	Y	-	-	-	Y	Y	-	0.78441	0.10224	0.18130	0.54764
14	Y	Y	Y	-	-	-	Y	-	Y	0.79130	0.14365	0.24309	0.56649
15	Y	Y	Y	-	-	-	-	Y	Y	0.78493	0.10664	0.18793	0.54949
16	Y	Y	-	Y	Y	Y	-	-	-	0.79325	0.22583	0.33784	0.59626
17	Y	Y	-	Y	Y	-	Y	-	-	0.79708	0.19265	0.30716	0.58726
18	Y	Y	-	Y	Y	-	-	Y	-	0.78778	0.13012	0.22264	0.55952
19	Y	Y	-	Y	Y	-	-	-	Y	0.79663	0.18688	0.30009	0.58496
20	Y	Y	-	Y	-	Y	Y	-	-	0.79618	0.21487	0.32986	0.59436
21	Y	Y	-	Y	-	Y	-	Y	-	0.78583	0.14955	0.24581	0.56497
22	Y	Y	-	Y	-	Y	-	-	Y	0.79700	0.21064	0.32641	0.59341
23	Y	Y	-	Y	-	-	Y	Y	-	0.78906	0.13326	0.22773	0.56143
24	Y	Y	-	Y	-	-	Y	-	Y	0.79798	0.18780	0.30260	0.58616
25	Y	Y	-	Y	-	-	-	Y	Y	0.78883	0.13606	0.23104	0.56225
26	Y	Y	-	-	Y	Y	Y	-	-	0.78816	0.28182	0.38386	0.61243
27	Y	Y	-	-	Y	Y	-	Y	-	0.78613	0.15219	0.24931	0.56605
28	Y	Y	-	-	Y	Y	-	-	Y	0.80060	0.23375	0.35390	0.60380
29	Y	Y	-	-	Y	-	Y	Y	-	0.78883	0.13509	0.23000	0.56191
30	Y	Y	-	-	Y	-	Y	-	Y	0.80000	0.20862	0.32768	0.59471
31	Y	Y	-	-	Y	-	-	Y	Y	0.78846	0.13632	0.23129	0.56209
32	Y	Y	-	-	-	Y	Y	Y	-	0.78718	0.15726	0.25624	0.56849
33	Y	Y	-	-	-	Y	Y	-	Y	0.80060	0.23489	0.35508	0.60422
34	Y	Y	-	-	-	Y	-	Y	Y	0.78666	0.15793	0.25675	0.56838
35	Y	Y	-	-	-	-	Y	Y	Y	0.78958	0.14356	0.24144	0.56532
36	Y	-	Y	Y	Y	Y	-	-	-	0.79025	0.20092	0.30950	0.58572
37	Y	-	Y	Y	Y	-	Y	-	-	0.79423	0.17530	0.28477	0.57942
38	Y	-	Y	Y	Y	-	-	Y	-	0.78591	0.11450	0.20013	0.55294
39	Y	-	Y	Y	Y	-	-	-	Y	0.79445	0.16795	0.27637	0.57701
40	Y	-	Y	Y	-	Y	Y	-	-	0.79280	0.18808	0.29796	0.58292
41	Y	-	Y	Y	-	Y	-	Y	-	0.78216	0.12557	0.21231	0.55431
42	Y	-	Y	Y	-	Y	-	-	Y	0.79400	0.18666	0.29741	0.58320
43	Y	-	Y	Y	-	-	Y	Y	-	0.78666	0.11247	0.19770	0.55271
44	Y	-	Y	Y	-	-	Y	-	Y	0.79580	0.16766	0.27726	0.57781
45	Y	-	Y	Y	-	-	-	Y	Y	0.78606	0.11270	0.19773	0.55238
46	Y	-	Y	-	Y	Y	Y	-	-	0.78621	0.26203	0.36456	0.60434
47	Y	-	Y	-	Y	Y	-	Y	-	0.78261	0.12442	0.21131	0.55422
48	Y	-	Y	-	Y	Y	-	-	Y	0.79790	0.21306	0.33007	0.59493
49	Y	-	Y	-	Y	-	Y	Y	-	0.78636	0.11214	0.19705	0.55240
50	Y	-	Y	-	Y	-	Y	-	Y	0.79970	0.19454	0.31204	0.58968
51	Y	-	Y	-	Y	-	-	Y	Y	0.78628	0.11180	0.19671	0.55223
52	Y	-	Y	-	-	Y	Y	Y	-	0.78328	0.12227	0.20864	0.55388
53	Y	-	Y	-	-	Y	Y	-	Y	0.79963	0.21857	0.33771	0.59798
54	Y	-	Y	-	-	Y	-	Y	Y	0.78321	0.12750	0.21551	0.55562
55	Y	-	Y	-	-	-	Y	Y	Y	0.78628	0.11354	0.19897	0.55281
56	Y	-	-	Y	Y	Y	Y	-	-	0.78486	0.29282	0.38925	0.61415
57	Y	-	-	Y	Y	Y	-	Y	-	0.78883	0.20500	0.31208	0.58619
58	Y	-	-	Y	Y	Y	-	-	Y	0.79775	0.25387	0.36978	0.60897
59	Y	-	-	Y	Y	-	Y	Y	-	0.79558	0.17948	0.29107	0.58176
60	Y	-	-	Y	Y	-	Y	-	Y	0.80127	0.22176	0.34281	0.60016
61	Y	-	-	Y	Y	-	-	Y	Y	0.79505	0.17146	0.28092	0.57863
62	Y	-	-	Y	-	Y	Y	Y	-	0.79243	0.19442	0.30454	0.58487
63	Y	-	-	Y	-	Y	Y	-	Y	0.80090	0.24696	0.36714	0.60865
64	Y	-	-	Y	-	Y	-	Y	Y	0.79385	0.19143	0.30217	0.58474
65	Y	-	-	Y	-	-	Y	Y	Y	0.79625	0.17430	0.28541	0.58040
66	Y	-	-	-	Y	Y	Y	Y	-	0.78478	0.26464	0.36538	0.60432
67	Y	-	-	-	Y	Y	Y	-	Y	0.78838	0.29217	0.39278	0.61623
68	Y	-	-	-	Y	Y	-	Y	Y	0.79685	0.21748	0.33329	0.59577
69	Y	-	-	-	Y	-	Y	Y	Y	0.79948	0.19780	0.31545	0.59068
70	Y	-	-	-	-	Y	Y	Y	Y	0.79850	0.22369	0.34144	0.59902
71	-	Y	Y	Y	Y	Y	-	-	-	0.79363	0.22624	0.33876	0.59666
72	-	Y	Y	Y	Y	-	Y	-	-	0.79693	0.19622	0.31103	0.58841
73	-	Y	Y	Y	Y	-	-	Y	-	0.79108	0.14846	0.24933	0.56805
74	-	Y	Y	Y	Y	-	-	-	Y	0.79700	0.19209	0.30649	0.58703
75	-	Y	Y	Y	-	Y	Y	-	-	0.79693	0.21790	0.33393	0.59592
76	-	Y	Y	Y	-	Y	-	Y	-	0.79115	0.16508	0.26971	0.57386
77	-	Y	Y	Y	-	Y	-	-	Y	0.79768	0.21614	0.33284	0.59576
78	-	Y	Y	Y	-	-	Y	Y	-	0.79205	0.14518	0.24587	0.56754
79	-	Y	Y	Y	-	-	Y	-	Y	0.79888	0.19077	0.30685	0.58779
80	-	Y	Y	Y	-	-	-	Y	Y	0.79190	0.14834	0.24974	0.56853
81	-	Y	Y	-	Y	Y	Y	-	-	0.78853	0.28482	0.38679	0.61374
82	-	Y	Y	-	Y	Y	-	Y	-	0.79130	0.17002	0.27571	0.57566

Continuation of Table

	LR	DT	SVM	ANN	KNN	NB	RF	ADA	GB	Accuracy	Recall	F1 score	ROCauc
83	-	Y	Y	-	Y	Y	-	-	Y	0.80097	0.23969	0.36019	0.60614
84	-	Y	Y	-	Y	-	Y	Y	-	0.79198	0.14893	0.25065	0.56878
85	-	Y	Y	-	Y	-	Y	-	Y	0.80135	0.21359	0.33451	0.59734
86	-	Y	Y	-	Y	-	-	Y	Y	0.79175	0.14984	0.25165	0.56894
87	-	Y	Y	-	-	Y	Y	Y	-	0.79220	0.16798	0.27391	0.57552
88	-	Y	Y	-	-	Y	Y	-	Y	0.80165	0.23982	0.36112	0.60664
89	-	Y	Y	-	-	Y	-	Y	Y	0.79235	0.17024	0.27693	0.57640
90	-	Y	Y	-	-	-	Y	Y	Y	0.79220	0.15293	0.25575	0.57029
91	-	Y	-	Y	Y	Y	Y	-	-	0.78673	0.31178	0.40646	0.62191
92	-	Y	-	Y	Y	Y	-	Y	-	0.79475	0.23193	0.34539	0.59935
93	-	Y	-	Y	Y	Y	-	-	Y	0.79970	0.27281	0.38902	0.61676
94	-	Y	-	Y	Y	-	Y	Y	-	0.79715	0.19780	0.31298	0.58911
95	-	Y	-	Y	Y	-	Y	-	Y	0.80232	0.23629	0.35844	0.60586
96	-	Y	-	Y	Y	-	-	Y	Y	0.79805	0.19589	0.31161	0.58903
97	-	Y	-	Y	-	Y	Y	Y	-	0.79715	0.22013	0.33642	0.59684
98	-	Y	-	Y	-	Y	Y	-	Y	0.80127	0.26312	0.38250	0.61446
99	-	Y	-	Y	-	Y	-	Y	Y	0.79813	0.21400	0.33107	0.59533
100	-	Y	-	Y	-	-	Y	Y	Y	0.79873	0.19488	0.31105	0.58912
101	-	Y	-	-	Y	Y	Y	Y	-	0.78868	0.28674	0.38848	0.61450
102	-	Y	-	-	Y	Y	Y	-	Y	0.78943	0.30892	0.40730	0.62269
103	-	Y	-	-	Y	Y	-	Y	Y	0.80127	0.24069	0.36140	0.60668
104	-	Y	-	-	Y	-	Y	Y	Y	0.80105	0.21577	0.33622	0.59790
105	-	Y	-	-	-	Y	Y	Y	Y	0.80127	0.23950	0.36018	0.60629
106	-	-	Y	Y	Y	Y	Y	-	-	0.78553	0.29480	0.39172	0.61529
107	-	-	Y	Y	Y	Y	-	Y	-	0.79265	0.21008	0.32135	0.59045
108	-	-	Y	Y	Y	Y	-	-	Y	0.79820	0.25457	0.37092	0.60952
109	-	-	Y	Y	Y	-	Y	Y	-	0.79640	0.18585	0.29906	0.58450
110	-	-	Y	Y	Y	-	Y	-	Y	0.80105	0.22369	0.34455	0.60068
111	-	-	Y	Y	Y	-	-	Y	Y	0.79603	0.18006	0.29185	0.58224
112	-	-	Y	Y	-	Y	Y	Y	-	0.79610	0.19992	0.31438	0.58919
113	-	-	Y	Y	-	Y	Y	-	Y	0.80135	0.25023	0.37073	0.61009
114	-	-	Y	Y	-	Y	-	Y	Y	0.79685	0.19814	0.31262	0.58904
115	-	-	Y	Y	-	-	Y	Y	Y	0.79723	0.17850	0.29124	0.58249
116	-	-	Y	-	Y	Y	Y	Y	-	0.78793	0.27071	0.37409	0.60849
117	-	-	Y	-	Y	Y	Y	-	Y	0.78838	0.29509	0.39510	0.61725
118	-	-	Y	-	Y	Y	-	Y	Y	0.79985	0.22524	0.34447	0.60044
119	-	-	Y	-	Y	-	Y	Y	Y	0.80075	0.20361	0.32312	0.59354
120	-	-	Y	-	-	Y	Y	Y	Y	0.80165	0.22825	0.34958	0.60268
121	-	-	-	Y	Y	Y	Y	Y	-	0.78561	0.29752	0.39393	0.61625
122	-	-	-	Y	Y	Y	Y	-	Y	0.78636	0.31834	0.41110	0.62397
123	-	-	-	Y	Y	Y	-	Y	Y	0.79888	0.25957	0.37609	0.61168
124	-	-	-	Y	Y	-	Y	Y	Y	0.80157	0.22713	0.34855	0.60221
125	-	-	-	Y	-	Y	Y	Y	Y	0.80210	0.25211	0.37330	0.61122
126	-	-	-	-	Y	Y	Y	Y	Y	0.78883	0.29785	0.39781	0.61849

End of Table

Table A.16: Scores of ensembles of size five - Korea

A.5 Hyperparameter Settings

Classifer	Duke	UCL	Chile	Korea
LogisticRegression	C=1000,	C= 0.1,	C= 1,	C= 100,
	solver = "sag",	max_iter=500,	max_iter=500,	max_iter= 500,
	max_iter = 5000	solver= 'saga'	solver='newton-cg'	solver= 'newton-cg'
DecisionTreeClassifier	criterion ="gini",	criterion ="gini",	criterion="gini,	criterion ="gini",
	max_depth = 5,	max_depth = 9,	max_depth = 8,	max_depth = 9,
	min_samples_split=38	min_samples_split=3	min_samples_split=3	min_samples_split=3
SVC	C=400,	C= 1000,	C=1000,	C= 100,
	kernel = "linear",	kernel = 'rbf',	kernel = 'rbf',	kernel = 'rbf',
	gamma='scale'	gamma='scale'	gamma='scale'	gamma='scale'
MLPClassifier	alpha= 0.1,	alpha= 0.1,	alpha= 0.001,	alpha= 0.1,
	hidden_layer_size=(11,11,11),	hidden_layer_size=(16,16,16),	hidden_layer_sizes=(20,20,20),	hidden_layer_sizes=(50,50,50),
	max_iter= 1000,	max_iter= 1000,	max_iter= 1000,	max_iter= 1000,
	solver= 'lbfgs'	solver= 'lbfgs'	solver= 'lbfgs'	solver= 'lbfgs'
KNeighborsClassifier	metric='minkowski,	metric='manhattan',	metric='manhattn,	metric= 'minkowski',
	n_neighbors= 6,	n_neighbors= 6,	n_neighbors= 8,	n_neighbors= 6,
	weights= 'distance'	weights= 'distance'	weights='distance'	weights= 'distance'
GaussianNB	-	-	-	-
RandomForestClassifier	max_features='sqrt',	max_features='sqrt',	max_features='auto',	max_features='sqrt',
	n_estimators= 700	n_estimators= 700	n_estimators= 200	n_estimators= 700
AdaBoostClassifier	n_estimators = 100	n_estimators = 100,	n_estimators=700,	n_estimators=700,
		learning_rate=1.0	learning_rate=0.1	learning_rate=0.1
GradientBoostingClassifier	n_estimators=100,	n_estimators=700,	n_estimators=700,	n_estimators=700,
	learning_rate=1.0,	learning_rate=0.05,	learning_rate=0.1,	learning_rate=0.05,
	max_depth=1	max_depth=4	max_depth=4	max_depth=5

Table A.17: Hyperparameter settings

References

Abbasimehr, H., Setak, M., & Tarokh, M. J. (2014). A comparative assessment of the performance of ensemble learning in customer churn prediction. *International Arab Journal of Information Technology*.

Basiri, J., Taghiyareh, F., & Moshiri, B. (2010). A hybrid approach to predict churn. In *Proceedings - 2010 ieee asia-pacific services computing conference, apscc 2010.* doi: 10.1109/APSCC.2010.87

Berger, H., Merkl, D., & Dittenbach, M. (2006). Exploiting partial decision trees for feature subset selection in e-mail categorization. In *Proceedings of the 2006 acm symposium on applied computing* (pp. 1105–1109).

Blockeel, H. (2018). Machine learning and inductive inference. *KU Leuven*.

Chali, Y., Hasan, S. A., & Mojahid, M. (2014). Complex question answering: Homogeneous or heterogeneous, which ensemble is better? In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics).* doi: 10.1007/978-3-319-07983-7_21

Cruz, R. M., Sabourin, R., & Cavalcanti, G. D. (2018). Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, *41*, 195–216. Retrieved from `http://dx.doi.org/10.1016/j.inffus.2017.09.010` doi: 10.1016/j.inffus.2017.09.010

De Bock, K. W., & Poel, D. V. D. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2011.04.007

De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*. doi: 10.1016/j.ejor.2018.02.009

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1–15).

Gallo, A. (2014). The value of keeping the right customers. *Harvard business review*, *29*.

Gashler, M., Giraud-Carrier, C., & Martinez, T. (2008). Decision Tree Ensemble: Small Heterogeneous Is Better Than Large Homogeneous. In *2008 seventh international conference on machine learning and applications.* doi: 10.1109/ICMLA.2008.154

Geeta, D. V., & Shashi, M. (2012). Applicability of Data Mining Techniques for Churn Prediction. , *VIII*(1), 22–36.

Gilpin, S. a., & Dunlavy, D. M. (2009). Relationships Between Accuracy and Diversity in Heterogeneous Ensemble Classifiers.

Huang, Y., & Kechadi, T. (2013). An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2013.04.020

Keramati, A., Aliannejadi, M., Ahmadian, I., Mozzafari, M., & Abbasi, U. (2014). Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing Journal*. Retrieved from `http://dx.doi.org/10.1016/j.asoc.2014.08.041` doi: 10.1016/j.asoc.2014.08.041

Layton, R. (2015). *Learning data mining with python.* Packt Publishing Ltd.

Lee, J., & Lee, J. (2006). Customer churn prediction by hybrid model. In *Advanced data mining and applications.* doi: 10.1007/11811305_104

Liu, Z., Dai, Q., & Liu, N. (2014). Ensemble selection by GRASP. *Applied Intelligence*. doi: 10.1007/s10489-013-0510-0

Mahajan, V., Misra, R., & Mahajan, R. (2015). Review of data mining techniques for churn prediction in telecom. *Journal of Information and Organizational Sciences*, *39*(2), 183–197.

Polikar, R. (2006). Ensemble Based Systems in Decision Making. *Ieee Circuits And Systems Magazine*, 21–45.

Sabzevari, M., Martínez-Muñoz, G., & Suárez, A. (2018). Pooling homogeneous ensembles to build heterogeneous ensembles. *arXiv preprint arXiv:1802.07877*.

Ting, K. M., & Witten, I. H. (1999). Issues in stacked generalization. *Journal of artificial intelligence research*, *10*, 271–289.

Tsai, C.-F., & Lu, Y.-H. (2010). Data mining techniques in customer churn prediction. *Recent Patents on Computer Science*, *3*(1), 28–32. doi: 10.2174/1874479611003010028

Tsoumakas, G., Partalas, I., & Vlahavas, I. (2009). An ensemble pruning primer. In *Applications of supervised and unsupervised ensemble methods* (pp. 1–13). Springer.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). Simulation Modelling Practice and Theory A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1–9. Retrieved from `http://dx.doi.org/10.1016/j.simpat.2015.03.003` doi: 10.1016/j.simpat.2015.03.003

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, *218*(1), 211–229.

Verbraken, T., Verbeke, W., & Baesens, B. (2013). A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE Transactions on Knowledge and Data Engineering*. doi: 10.1109/TKDE.2012.50

Whalen, S., & Pandey, G. (2013). A Comparative Analysis of Ensemble Classifiers : Case Studies in Genomics. In *Proceedings - ieee international conference on data mining, icdm (2013) 807-816.*

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization.

IEEE transactions on evolutionary computation, *1*(1), 67–82.

Zhang, Y., Qi, J., Shu, H., & Cao, J. (2007). A hybrid KNN-LR classifier and its application in customer churn prediction. In *Conference proceedings - ieee international conference on systems, man and cybernetics.* doi: 10.1109/ICSMC.2007.4414197

FACULTY OF BUSINESS AND ECONOMICS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be

KU LEUVEN
FACULTY OF ECONOMICS AND BUSINESS

Optimal mix for selective heterogeneous ensemble learners for churn prediction in the telecommunication industry

Individual Report

Raphael Hanke (0736510)

Thesis submitted to obtain the degree:

Master of Science
Information Management

Promotor: Prof. Dr. Vanthienen
Supervisor: Ziboud van Veldhoven

Academic year: 2018/2019

Chapter 1

Individual Report

Parties Involved
The team engaged in writing this thesis consisted of Koen Barbiers, Stefan Hegedis and Me, Raphael Hanke. We got to know each other at the thesis information session and formed the group based on similar interests. Each of us had strong interests in a data driven research topics related to the application of machine learning and analytics in the business environment. After having written the proposal, we were satisfied to obtain our first priority pick. We were glad to cooperate with Prof. Dr. Jan Vanthienen as coordinator and Ziboud van Veldhoven as direct supervisor, who gave us great support and assistance the following months.

Time	Activity
05.10.18	Group formation and topic motivation
29.19.18	Plagiarism test
November & December	Literature study
30.11.18	Literature review draft
03.12.18	First seminar
January	Break for exam period
February	Methodology
February	Data preprocessing & first models
18.02.19	Mid-term presentation
March	Run scripts for ensembles
April	Evaluation & processing of outputs
April	Conclusion & finalize report
19.04.19	Final draft
May	Integrate feedback & adjust formatting
16.05.19	Final report deadline
29.05.19	Thesis defense

Table 1.1: Timeline of the master project

Chronological Overview
The timeline in table 1.1 provides a broad overview of the milestones, deadlines and activities which were connected to the master project and guided the work. Throughout the semester we held personal meetings with promoter and daily supervisor. Moreover,

we agreed to keep our supervisor up to date with our progress via mail to every two weeks. Group meetings were scheduled at least once every week, depending on the progress and upcoming deliverables. In our meetings, we discussed the progress and assigned tasks to finish until the next one. Besides of that, we have seen each other daily due to overlapping class schedules and thus always had opportunities to talk about current issues. During the meetings, as well as the official events, the group has always been complete.

Individual Reflection and Vision

Since the acquisition of customers costs up to six times as much as customer retention, the relevance of churn prediction in the telecommunication industry is undeniable (Gallo, 2014). Due to this cost sensitivity, performance improvements of churn prediction models are beneficial for firms. Ensemble methods are known to achieve a higher performance than single classifiers. However, the complexity of ensembles leads to individual choices that are connected to their construction. Our research about the optimal mix for heterogeneous ensemble learner aimed to ease model construction with a focus on important parameters, such as the base classifier selection, the number of components and powerful classifier combinations.

Intensive study of literature unveiled many opportunities. Determining the own position within in a very complex research field has been challenging. An infinite pool of base classifiers, different aggregation methods and many newly developed dynamic and static classifier selection methods gave a lot of room for potential research areas. On the other hand, this also highlighted the necessity for limiting ourselves. We decided to build all possible ensembles in order to analyze and compare composition with performance. We chose a rather pragmatic approach for the technical implementation. By setting constraints in a) the number of base classifiers, b) the ensemble size, c) potential base classifiers and d) aggregation method, we were able to reduce the complexity of the problem. The creation of all possible ensembles within those boundaries nevertheless promised interesting results.

All possible ensembles in size three and five of 11 base classifier led to 210 models per dataset, whose performance was assessed. In comparison to existing studies this was a new approach, as they usually focus on the evaluation of a specific ensemble. Our results provide recommendations for the ensemble construction phase as well as an optimal mix of base classifier.

Individual Suggestions for Improvement

Loosening constraints gives clear perspectives for potential extensions of the research. Further research in a similar fashion would be possible with ensembles of size bigger than five, with an extended pool of base classifiers or other aggregation methods such as stacking. Also the field of application could be widened beyond churn prediction. However, less constraints are connected with an increased complexity. Alternative performance measures, next to the F1-score, could be added to the analysis. Do results based other evaluation metrics confirm our findings? Ensemble selection methods are an important subtopic within the literature of ensemble methods with a lot of recent development and new publications about i.e. dynamic selection methods (Cruz, Sabourin, & Cavalcanti, 2018). This research was conducted independently of any base classifier selection method. A comparison of between the selection-method-based results and the results of this research would be an interesting extension.

Self-Evaluation and Team-Evaluation

We decided not to divide the work upfront into different parts. By doing so each of us was involved and contributed to every step of the thesis. On a weekly base it was necessary to assign tasks in order to manage the project efficiently. Stefan focused i.e. more on the data preprocessing and parameter tuning. Koen and me worked on evaluating the results and writing the final report. We made use of Mendeley to share, collect and organize literature. Git and SourceTree were used to share code by working on common repositories. Overleaf and Google Docs helped us to write the report together. By doing so, we were able to effectively communicate, review each other's work and suggest improvements.

Working in a team with Koen and Stefan was very pleasant. In retrospective, I am happy that we found each other. As an international team with members of different academic backgrounds, we were able to complement each other with individual qualifications. Different ways of thinking and approaches on solving problems led to constructive discussions. Both Koen and Stefan did a great job.

References

Cruz, R. M., Sabourin, R., & Cavalcanti, G. D. (2018). Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, *41*, 195–216. Retrieved from http://dx.doi.org/10.1016/j.inffus.2017.09.010 doi: 10.1016/j.inffus.2017.09.010

Gallo, A. (2014). The value of keeping the right customers. *Harvard business review*, *29*.

FACULTY OF BUSINESS AND ECONOMICS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be