Two class Bayes point machines in repayment prediction of low credit borrowers

David Maloney∗, Sung-Chul Hong, Barin N. Nag

Information Technology, Department of Computer and Information Sciences, Department of Business Analytics & Technology Management, Towson University, Towson, MD 21252, United States of America

A R T I C L E   I N F O

Dataset link: https://www.kaggle.com/eho0426/lending-club-2017-2020

Keywords:
Machine learning
Risk analysis
Decision analysis
Bayesian classification
Prediction

A B S T R A C T

Decision-making in the peer-to-peer loan market has not been studied as extensively as traditional lending mostly because of the perceived risk in dealing with low credit borrowers seeking funding alternatives. We develop a machine learning-based approach to test the viability and usefulness in peer-to-peer loan repayment predictions among low credit borrowers. This analysis provides potential benefits that could strengthen the lending market with a more reliable method of identifying applications from promising candidates with low credit. Here an experiment will be performed to measure the performance of a model used for classifying peer-to-peer loan data. The aim is to aid the repayment prediction capabilities of peer lenders when analyzing low credit applicants. A binary classification algorithm is used to build the model and applied to actual historical loan data to evaluate performance. Experiment results, visualizations, and key performance indicators are discussed in the work to influence confidence in using the method proposed.

1. Introduction

Peer-to-peer lending is the concept of individuals intentionally selecting and lending funds to other individuals (or small businesses). There are a few larger online platforms that pair individual lenders with individual borrowers. Many other smaller players have entered the space fulfilling a need that predatory lending practices have previously filled. Before peer-to-peer lending became as accessible and reputable, the alternative lending market was dominated by the payday loan and title loan businesses. Lending outlets like these are very accessible to borrowers but set unreasonable terms for repayment. With the increasing adoption of various aspects of fintech came the ability for individuals to transact with one another without physical barriers (Liu et al., 2020). Lending naturally found its place as a subset of the multiple types of transactions possible within the realm of technology-enhanced financial innovation.

The general idea is that peer-to-peer lending is another option that can be utilized when the traditional lending construct either does not give favorable terms or is not available. Traditionally, larger institutions make decisions and lend to applicants based mainly on credit profiles. Replacing the large institution with an individual allows that individual or peer to become the lender and ultimately the decision-maker. This can yield varying results as the peer may decide to either stay consistent, tighten, or relax the criteria the borrower must meet to qualify for the same loan at the institutional level. The advantage is that there are likely to be multiple potential lenders competing to make this decision (Liu et al., 2020). The criteria for decision-making can change when the qualifying requirements are less rigid because there will be different risk profiles associated with each lending peer’s preference (Boiko Ferreira et al., 2017). So, a borrower with lower credit can typically qualify for loans in the peer network because of an alignment with a corresponding lending tier or grade that matches the applicant’s profile ( Fitzpatrick and Mues, 2021).

The downside to offering a wider range of criteria is that risk has historically only been based on credit, which would mean there would be a higher risk of default. The problem with this is that an alternative lending method that does not also use alternative risk assessment to qualify might just as well be considered subprime lending when dealing with low credit applicants. To offer a change, a difference in risk assessment such as a different assignment of weight for certain attributes in the applicant profile could give some lower credit individuals a better chance at qualifying. The common issue with any lending format derivation would continue to be the same since the investment is not guaranteed due to the possibility of borrower default. Peer lenders taking risks on

∗ Corresponding author.
E-mail address: damalon6@students.towson.edu (D. Maloney).

https://doi.org/10.1016/j.heliyon.2022.e11479
Received 11 June 2022; Received in revised form 22 August 2022; Accepted 2 November 2022

2405-8440/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
lower-tier applications are rewarded with higher interest. Lenders must also be mindful enough to analyze whether the burden of interest gives the borrower a better probability of repayment as full repayment should be prioritized more than being able to charge higher rates. Rates can be set and agreed upon by the borrowing and lending peer resulting in a contract with a rate and term for repayment. Lending peers are responsible for analyzing the attributes of the applicant profile to determine suitability (Guo et al., 2016). Using alternative attributes and methods for analysis of loan applications has produced interesting results in work done by others using digital footprint and text-based data of the applicant (Kriebel and Stitz, 2022). Non-traditional methods like these should be explored more in alternative lending to truly provide an alternative lending experience. Since alternative prediction methods aren’t widely adopted, the peer-to-peer space has become more of a low tiered lending experience differentiated by the source of funding and higher rates. Investors can set their risk level in notes they intend to expose their funds to on the marketplace. In this regard, some methods of note selection could prove to be more beneficial in risk avoidance. Analysis of historical repayment performance using machine learning techniques can give peers information to make good decisions on funding requests.

By applying machine learning techniques to the data offered on these platforms, the decision-making capabilities are supported by data-driven predictions from actual data instead of speculation. This is a change from the traditional lending construct. The individual becomes the direct lender and can make decisions with prior loan data. Numerous efforts have been made to test the decision-making capabilities of individuals based on having prior information. Bayesian tracking algorithms can be used for effective methods in reducing bias and improving performance by limiting misclassifications (Short and Mohler, 2022). The importance of Bayes-revised probabilities on decision makers could pose a problem to some that want to take ownership of their decision process, but in the case of the individual that prefers the assistance, Bayesian classification can bridge the gap (Goodwin et al., 2018). The peer lender stands to gain an advantage if there is a consistent repeatable level of accuracy in the prediction capabilities when applying machine learning techniques to the loan performance data offered. In this work, the prediction capability will be analyzed to determine if consistent enough to be used by the peer lender in making confident decisions on applications from individuals with low credit.

All listings on the marketplace are posted by borrowers requesting loans. As shown in Table 1, a favorable score for lending using this metric only would be 670 or greater. The targeted group for this research experiment will have “Fair” and “Very Bad” ratings based on the Score range in Table 1. The lending friction as described in the Impact column of Table 1 becomes evident when the score drops below 670 as the applicant may not secure funding. The first objective of the study is to examine and understand some of the factors that play a role in the low credit borrowers represented in this portion of the peer-to-peer loan dataset. In that segment of borrowers, we intend to analyze the repayment performance of these individuals with lower credit scores. We will also use machine learning to train a model for predicting repayment success specifically in this group. The purpose and second objective is to determine whether predictions in repayment can be consistently performed. A resulting model will be built using the factors discovered in the first objective. The performance will support either recommending or discouraging peer lenders from utilizing this method. From the results of the experiment individual peer lenders will have the information needed to determine if the Two Class Bayes Point Machine is a viable method for the selection of low credit applications. One potential benefit of finding success in this method would be an influence to increase the number of low credit applications approved and repaid in the peer lending market. Fig. 1 shows the difference in lender confidence in repayment among the various credit groups. Lenders would prefer to lend to creditworthy borrowers, but there is evidence of a tendency to repay among the low credit group as shown in Fig. 1. Confidence is important when identifying applications in the low credit segment because this alternative lending space exists to provide mutual benefit to borrowers and lenders. This benefit can only be achieved when loans are approved and repaid successfully. An increase in the ability to identify promising low credit loan applicants can encourage more inclusion of this low credit group in peer lending. In short, the proposed model answers two research questions:

a) Which attributes of loan applications show to have impact on repayment predictions in the low credit group?

b) Can the Two Class Bayes Point Machine be used to make predictions in the low credit group that will perform better than the baseline?

This paper is organized in the following manner: Section 2 contains related work in loan repayment prediction. In section 3, there is an explanation of the machine learning algorithm chosen to do this work. We further discuss the machine learning application, why it applies to the problem, and its performance in predicting repayment. Section 4 will give an explanation of the dataset, the attributes used, and how these features were selected. In section 5 the experiment details are presented with results. The results of the findings are discussed in section 6, and the work is concluded in section 7.

2. Related work

Repayment predictions have been studied extensively in various research. A summary of some of the contributions to this space and how
they differ from this work can be found in Table 2. Ultimately the prevailing model in the U.S. uses the credit score for determining the ability to repay (Berg et al., 2020). Many institutions and lending organizations use the credit report and score as the basis for examining an individual’s previous financial interactions involving a lender that reported performance to any of the three major bureaus. Predicting loan repayment has been studied extensively. Kim and Cho discussed a promising model built using a convolutional neural network (CNN) to predict repayment in peer-to-peer lending with 5-fold cross-validation automatically pulling out complex features more effectively than many other machine learning methods for prediction (Kim and Cho, 2019). They were able to develop empirical rules for applying neural network algorithms to data like this and efficiently predict patterns pointing to a potential default of borrowers. One beneficial contribution made in their work is that no features were extracted in their analysis, which proved beneficial to developing the multiple layers of the network. This is important because there are many attributes in the data that are seemingly useless and are often removed from the dataset by many researchers who use peer-to-peer lending data (Kim and Cho, 2019). Our work can provide an extension by utilizing this feature selection process on specific segments of the total population of borrowers. Since the authors used a randomized subset as the basis of their experiment, their model can provide a great starting point for making predictions in a specified group such as those with low credit.

Bayesian analysis is commonly used in credit analysis and repayment prediction. A Bayesian competing risk hazard model was developed in the work done by Bhattacharya et al. where default and repayment data were used with Markov Chain Monte Carlo methods (Bhattacharya et al., 2019). Though the work was based on the duration of single home mortgages instead of personal loans, the model similarly analyzed statistical differences in default and repayment probabilities. A key difference in their work was the analysis of a third outcome defined as “prepayment” which they called a competing risk in such a way that the model predicts which would be likely to occur first. One interesting aspect of their findings was that the rate of default and the rate of prepayment behaved identically for the debt-to-income, and occupancy attributes. The difficulty experienced in estimating default times was due to a smaller ratio of default observations in the overall total dataset (Bhattacharya et al., 2019). The data used in that work is unbalanced, with 1% to 2.5% of the loans defaulted. This is a common issue that our model also experiences as the number of defaults are not equal in size to the number of successful repayments. Our research can extend this work by applying the predictive capabilities of the Two Class Bayes Point Machine among the low credit group to the subprime loan market. This contribution would be justified because many of the low credit borrowers in the subprime market default at a rate of over 14% in some years in comparison to the 2.5% rate observed in the dataset used (Bhattacharya et al., 2019).

In the work done by Lui et al. Bayesian analysis is used to determine the relationship between the probability and timing of default. Their work, using mixture models to represent the unidentified subpopulation, proved to be successful in identifying default time. A mixture of both hierarchical and non-hierarchical models were used to show its usefulness to banks to better understand the individual’s actions. Their findings are promising for providing timely capital risk management indicators for lenders (Liu et al., 2015). Their method utilized Markov Chain Monte Carlo omitting variable selection, time-based variables like the prime (variable) interest rate, and provided a useful Bayesian based system effective in default timing.

Chen et al. have done work in the peer-to-peer lending space, explicitly solving the issues arising from dealing with imbalanced datasets (Chen et al., 2021). This work, though very detailed, focused on a common problem among publicly available datasets in this space from the major platforms. This work made a significant contribution because it pointed out that analyzing performance data post-funding skews the dataset since only candidates with specific attributes received funding

Table 1. Credit Score Ranges, Ratings, and Impact on Lending (Source: Experian).

| Score Range | Rating   | Impact                                                                 |
|-------------|----------|----------------------------------------------------------------------|
| 300-579     | Very Bad | Credit applicants may be required to pay a fee or deposit, and applicants with this rating may not be approved for credit at all. |
| 580-669     | Fair     | Applicants with scores in this range are classified as subprime borrowers and may not be approved for credit by strict lenders. |
| 670-739     | Good     | Only 8% of applicants in this score range are likely to become seriously delinquent in the future. |
| 740-799     | Very Good| Applicants with scores here are likely to receive better than average rates from lenders. |
| 800-850     | Exceptional | Applicants with scores in this range are at the top of the list for the best rates from lenders. |

Table 2. Example of the Need for Analysis of Repayment Predictions among Low Credit Borrowers.

| Reference               | Method Used                  | Credit Scoring Data Set Analyzed | Analysis Specific to Low Credit Borrowers (Any Segmentation by Credit) |
|-------------------------|------------------------------|---------------------------------|---------------------------------------------------------------------|
| Kim and Cho (2019)      | Deep convolutional neural network (CNN) | Peer-to-peer loans Randomized subset 143,823 observations from 2015-2016 | No |
| Guo et al. (2016)       | Kernel regression            | Peer-to-peer loans Randomized subsets | No |
| Junior et al. (2020)    | kNN/RMkNN                    | Peer-to-peer loans Term of 36 months Q1, Q2, and Q3 | No |
| Zanin (2020)            | Regularized logistic regressions (ridge, lasso, and elastic-net) with class rebalancing | Peer-to-peer loans 2010 Q1 through 2015 Q4 Term of 36 months | No |
| Berg et al. (2020)      | Cramer’s V                   | Online purchase financing       | Yes – Separation of applicants with high credit score and applicants without an available credit score |
| Bhattacharya et al. (2019) | MCMC                        | Freddie Mac mortgage loans 1999-2014 | No |
while others were left out (Chen et al., 2021). Lenders are more likely to fund when there is a promising candidate presented. This vote of confidence is given based on a particular model making the recommendation. In this regard, the dataset becomes imbalanced because this will result in more loans with certain attributes as the applicants without these attributes are not funded (Chen et al., 2021). This affects the sampling rate of the data involved as the two possible outcomes are not equal in representation, resulting in skewed classification output. Their research addressed this problem by utilizing under-sampling and cost-sensitive learning to classify the imbalanced dataset.

Credit fairness has been analyzed from a machine learning perspective by others to maximize its application in preferable ways that limit monetary losses by lenders. Separation, which forces independence on a scoring model’s results, was found to be the most appropriate metric for credit scoring when comparing different fairness criteria (Kozodoi et al., 2022). This is because sensitive attributes such as those that coincide with historically unprivileged groups may correlate with default rates. The problem which persists across much of the work in this space is that even after the findings there is a group that does not meet the cut. What should be done, if anything, for the applicants who do not meet the credit cut off? Should there be a marketplace or alternative for the rejected lower credit applicant to turn to in respect to meeting lending needs. Separating aspects of the credit scoring method as the heaviest metric requires the underlying credit scoring model to continuously recommend against lending to those with unfavorable scores without digging into the factors weighing the score down. In one of the credit based analyses, the authors found that the default rate in the lowest credit score quintile was 2.12%, which was 5 times the default rate in the highest credit score quintile (0.39%), and more than twice the average default rate of 0.94% and Berg et al. (2020). In this work the authors also found that digital footprints did not seem to have an association or relationship with the borrower’s credit quality, and that digital footprint information should be used as a complement instead of a replacement for credit scoring. There was no correlation found between digital footprint data and credit scores among borrowers with credit info available (Berg et al., 2020). Consideration is due as the authors excluded applicants with low credit scores from the analysis. Among the remaining applicant data, credit score-based quintiles were established from highest to lowest. One way that our work can extend this research is the further analysis of the omitted low credit group to give a more accurate picture of the entire population. The exclusion of low credit applicants is in line with a typical credit card, bank loan, and peer-to-peer lending data set where low credit applicants are usually excluded from accessing credit (Berg et al., 2020).

Many have done work utilizing Bayesian analysis for predictions. Binary classification problems can be handled with fewer misclassifications when using linear programs that utilize Bayes Theorem. In Bayesian classification, the objective is to build data categorization logic that can be applied to records to assign a class based on the attributes. When employing Bayesian classification, probabilistic relationships are modeled based on the attributes. From the modeled relationships, an estimation of class membership can be made. When working with outcomes that are less critical in estimates in prediction, less sophisticated approaches are allowed, but when there is a large downside to an incorrect prediction, a more sophisticated process should be used (Gadzinski and Castello, 2020). Though conceptually simple, a key strength of Bayesian classification is exploiting the level of error probability in predictions.

Herbrich et al. proposed the Bayes Point Machine and proved that in the zero-training error case, this model provides exceptional performance. In their work, they explain at the heart of the Bayes Point Machine is a Bayes-optimal classifier that predicts the class that minimizes error most when applied to all possible combinations of the sample data. This is an expensive approach that can be time-consuming, especially in cases where there are millions of data points. To taper the cost, approximations can be made in applications where being precise is not as critical (Herbrich et al., 2001). In the loan repayment prediction use case, approximation is acceptable since there are many factors at play and generalization is applicable to the domain. Generalization in Bayesian classification is commonly achieved with initial randomization of the prediction capabilities of test data plots. A model can be constructed based on the hypothesis closest to prediction and then applied to a set of data points for evaluation. Another approach to generalization can be the use of a linear classifier that splits the data based on class and continuously classifies any new or incrementally introduced data points based on a hypothesis while evaluating accuracy (Abdelmoula, 2017).

Bayes Point Machines use a statistical approach to attempt to perfectly separate a training set so that a repeatable likelihood calculation can be applied to actual test data for accurate results. The algorithm itself of classification iterations that result in error while in the training phase. While helpful for accuracy on training data, this could lead to an overfitting side effect when exercised on test data. In the training phase, the output is a set of classifiers that operate within an operating range or space of version. The center of mass data plots in the operating range is optimized for classification capabilities based on the trained data provided. This area is the Bayes Point classifier for the dataset (Corston-Oliver et al., 2006).

3. Methodology

Type A predictions, or category forecasts, involve identifying one of two possible outcomes as the most probable. Such predictions can be processed visually with a tree like structure showing the attributes which led to the final decision. The loan applications provide multiple attributes and events such as income which can be represented by $a_{1}$ and $a_{2}$ in Fig. 2. These and other attributes of an observation, $C_{1}$ and $C_{2}$, are the possible events or attributes influenced by the attributes in the previous step. Let $j = P_{2}$ or the posterior probability using a Bayesian method (Goodwin et al., 2018). Though simplified in this example, with a tree like diagram shown in Fig. 2, Bayesian classification predictions can be followed back through the steps to show the logic that went into calculating the result.

In repayment prediction, the focus is on finding an effective method to correctly classify an observation of class $y$ which will in turn reveal feature set $x$ of which information can be used for evidence to classify more data in the domain $D$ and utilized in the training set (Herbrich et al., 1999).

$$D = \{y_{i}, x_{i}\}, i = 1, ..., N$$ (1)

Prediction requires applying $p(y = c \mid x, D), c = 1, ..., K$ over each of the two possible classes. The conditional factors in the formula (1) are $x$ and $D$ which we supply a description of the prediction $p(y \mid x, D)$ given a set of conditions (Abid et al., 2017). 3.1. Two class Bayes point machine for repayment prediction

The Two Class Bayes Point Machine, the algorithm used, works in the following manner. In (2), an assumption is made of a prior distri-
bution over \( w \). Every record then revises the belief of \( w \) and produces a posterior distribution. This posterior distribution is used to create the final \( w_{BPM} \) to determine class:

\[
w_{BPM} = E_{p(w|D)}[w] = \sum_{i=1}^{|V(D)|} p(w_i | D) w_i
\]  

(2)

Where \( p(w_i | D) \) is the posterior distribution of the weights given the fact the \( D \) and \( E_{p(w|D)} \) is assumed based on the distribution. \( |V(D)| \) is the size of \( V(D) \), or version space given the weights \( w_i \) which is consistent with the set of \( w_i \) that classifies with zero error on the training data. The Bayes optimal solution is the approximation based on the training set (Abdelloua et al., 2017). Once the heavy lifting of training the model is done, the model is then scored. Scoring of the model is necessary as it feeds the evaluation which will be analyzed further (Gadzinski and Castello, 2020).

For a set of \( 2 \) classes, \( a_i, \ i = 1, 2, \ldots, M \), the posterior probability that applies for each is \( p(a_i | x) \) to classify an observation, \( x \). Another approach constrains on the false-negative error and minimizes the false positive rate. This constrained optimization method can assume each member comes from a single population of input probabilities (Banks and Abad, 1991). The Bayes Point Machine, which is a probabilistic model for classification, uses a procedure to determine the class \( y \) to which an instance of interest belongs, given relevant feature values \( x \) of this instance and the information in \( D = \{ (y, x_i), \ i = 1, \ldots, N \} \) the training set of observed class labels and corresponding feature values.

For binary or two class label prediction \( y \in \{1, -1\} \) there are a few key aspects that must be accounted for. The first is that all features must be available for each observation. This is important as any missing features for any records will require a different approach for classification (Abdelloua, 2017). In preprocessing, missing values must be filled in, or the entire record will need to be excluded from the dataset. This missing data can potentially negatively affect the accuracy of the prediction capability. The next aspect is that each record has the same weight and level of importance. There must be an exchangeability to each instance such that \( p(y | x, w) \) has a predictive distribution that for any \( x \) the same parameters apply and records can be analyzed in any order. The prediction method must use a linear discriminant function of the form:

\[
p(y | x, w) = p(y | s = w^T x)
\]  

(3)

Where in (3), \( s \) is the score of the observation after applying the algorithm. To allow for errors in labeling, there should be Gaussian noise added to the score. This method of probit regression is necessary for handling measurement errors (Li et al., 2020).

\[
p(y | x, \epsilon) = 1(y s + \epsilon > 0) \quad \text{with}
\]

\[
p(\epsilon) = N(\epsilon | 0, 1) \quad \text{and} \quad 1(a > 0) = \begin{cases} 1 & \text{if } a > 0 \\ 0 & \text{otherwise} \end{cases}
\]  

(4)

This use of the Bayesian strategy in (4) allows for us to make assumptions about a probability distribution \( p \) over the binary set of outcomes. The probability models a belief about the likelihood of getting a different outcome (Abdallah et al., 2017). When dealing with outlier data, a method like the Gaussian prior distribution needs to be employed values that are not in the regular range of proximity to the mean of the weight distribution. To shift these feature values, bias must be introduced for all instances, but in such a way that avoids correlation.

The key difference in Bayes Point Machines and other classification methods that use Bayesian classification is the training of the weights (Compton-Oliver et al., 2006). For instance, another Bayesian classification method would classify a data point \( x \) based on \( y = (w^T x) \) for parameter vector \( w \) and \( y = \pm 1 \) in the linear classification scenario. Thus, given a training set \( D \)

\[
D = \{ (x_1, y_1), \ldots, (x_n, y_n) \}
\]  

(5)

the likelihood for \( w \) in can be written as:

\[
p(Y | w, X) = \prod_i p(y_i | x_i, w) = \prod_i \Psi(y_i, w^T \phi(x_i))
\]  

(6)

Where in (6), \( Y = \{y_i\}_{i=1}^n, X = \{x_i\}_{i=1}^n, \Psi(\cdot) \) is used for the cumulative distribution for a Gaussian and \( \phi(x_i) \) allows the classification region to take a nonlinear boundary form in the original features. To train the weights of a new input \( x_{n+1} \) the distribution for prediction is approximately:

\[
p(y_{n+1} | x_{n+1}, Y) = \int p(y_{n+1} | x_{n+1}, w) p(w | Y) dw \approx P(y_{n+1} | x_{n+1}, (w))
\]  

(7)

The \( w \) in (7) denotes the posterior mean of the weights and is also called the Bayes Point. The expected outcome for each observation is a Bayesian average which gives a generalization in class. The center of mass approximates the real Bayes Point (Herbrich et al., 2001). This algorithm was selected due to being a probabilistic model that makes probable predictions for new examples. Specifically, the Two Class Bayes Point Machine answers the question: What is the most probable classification of the new instance given the training data? This is seemingly in line with the task at hand of being able to predict repayments given the historic performance of other instances. This algorithm uses stochastic approximations to produce a center of mass with soft-boundaries to allow for training errors in noisy data. The Bayes Point Machine consistently outperforms the Support Vector Machine when using real data sets (Herbrich et al., 2001). For any given data point, the Bayes-optimal classifier predicts the class with the least average error when evaluating over all the boundaries and data samples possible. This is a gift and a curse as the more data in the sampling the less error, but the more possibilities and more expensive. Normally approximation is made by randomly plotting test points and evaluating the Bayes-optimal hypothesis prediction across all other points. The method is very slow for larger datasets, but to improve this a linear classifier is used. This linear approach essentially draws a line between the plot groupings of the two classes and classifies by plotting on either side of the boundary. The correctly classified center of mass for the training subset is used as the Bayes-optimal classifier approximation (Herbrich et al., 2001).

4. Data and attributes

The official historical data are offered online and are accessible by navigating to the statistics/data area for developers on the Lending Club website. These data sets are published quarterly as far back as 2015 and can be accessed much easier on data analysis platforms such as Kaggle. The files contain application information that depicts the borrower’s credit profile, repayment history, income, demographics, and use for the funds. The details of each loan in circulation are refreshed and published each quarter. Loan features such as loan amount, interest rate, repayment term, monthly payment amount, and loan status are included. Regarding the borrower, the file also contains attributes like annual income, employment, FICO score, homeownership, and the number of open accounts for credit. The list of attributes (features) shown in Table 3 with no particular order, is the standard attributes for lending decisions and credit-based business activity. Data then was properly formatted and assigned a data type in the system (numeric, categorical, string, date/time, or binary) to process in the next step. The dataset was separated into two random groups (one for training and one for testing) at various levels utilizing a randomized split function. Increments of 10 were used to capture the ratio of training to test data. Since there are only two classes, we employ an approach geared towards the prediction of outcomes limited to two classes based on given attributes in the data.

\[1\) Source: https://www.kaggle.com/ethon0426/lending-club-20072020q1.\]
This work is focused on the low credit portion of the population of borrowers. By filtering the dataset to have only a low credit population, essentially our results will determine which features have the most impact on loan repayment predictions within this group. We must analyze the relationship between the available attributes and repayment outcomes to make the case. On an individual feature basis, we test for the difference in mean, so $H_0$ = The attribute has no role to play in successful repayment of a peer-to-peer loan. In testing this, we find the meaningful attributes that play a role in successful prediction of repayment and can remove those that appear not to have an impact. A few of the key categorical variables along with their type and a sample of the values possible in the dataset are shown in Table 4. The output in our experiment is binary, categorical, and can only be default or repaid. Since the experiment utilizes a combination of categorical and numerical input, we must choose the technique that is best for removing non-informative predictors from the model by inspecting each feature separately. For the loan repayment prediction scenario, we use a supervised filter to make the model more interpretable and to make model training faster.

We employ Chi-Square to test correlation and determine relationship between the categorical variables and the categorical binary outcome. This test will allow for an examination and comparison in determining independence. The method attempts to maintain classification accuracy and original class distribution while selecting the minimal sized subset of features. Chi-Square which can be calculated with the formula (8) where $c$ = degrees of freedom, $O$ = observed value, and $E$ = expected value.

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

From the test performed on the dataset, the resulting P-values were used to determine what features should emerge to be used in the classification stage of the experiment. The test shows that a number of features show the possibility of having a statistical influence on the repayment outcome. Table 5 shows the categorical features that we must reject $H_0$ and their corresponding P-values. In the test the employment length, verification of application information, term of loan, housing situation, and loan purpose attributes all had P-values lower than the level of significance. These were the categorical attributes that were selected to be included in the experiment due to the calculated statistical significance.

We employ T-Test to compare the difference in group mean to determine relationship. Random sampling of continuous variables in our data will be compared with corresponding random samples of our categorical binary class to determine goodness of independence. The independent-sample T-test compared the means of continuous, independent variables in the two groups (Repaid = 1 and Repaid = 0). The T-Test, which can be calculated with the formula (9) where $\overline{x}_1$ = observed mean of 1st sample, $\overline{x}_2$ = observed mean of 2nd sample, $s_1$ = standard deviation of 1st sample, $s_2$ = standard deviation of 2nd sample, $n_1$ = size of 1st sample, and $n_2$ = size of 2nd sample.

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The test showed that there were statistically significant differences between the means of the group of people who repaid and those of the non-repaid group. Table 6 shows the continuous attributes that showed statistical importance. These attributes with P-values below the threshold of .05 were identified to be used in the experiment.

5. Experiment

The experiment is based on the cited dataset spanning from 2015 through and including 2019. The results of the model’s performance on this dataset, can be analyzed using multiple performance indicators for scoring the model. The accuracy, precision, recall, and F1 of the model will be used as measurements to score the classification capabilities of the model. The main classification problem in the experiment involves
Table 5: Resulting analysis of statistically important categorical features.

| Feature                  | Repaid Observed | Repaid Expected | Default Observed | Default Expected | $X^2$   | P-Value |
|--------------------------|-----------------|-----------------|------------------|------------------|--------|---------|
| Employment Length        | Not Working (Null) | 1151 | 1155 | 415 | 411 | 455.5 | .0001 |
| Unemployed (0)           |                 | 7297 | 8082 | 3664 | 2879 |
| < 1 (5)                  |                 | 8986 | 9203 | 3496 | 3279 |
| 1 year                   |                 | 8260 | 8426 | 3168 | 3002 |
| 2 years                  |                 | 10343 | 10344 | 3686 | 3685 |
| 3 years                  |                 | 10085 | 10048 | 3543 | 3580 |
| 4 years                  |                 | 7625 | 7676 | 2785 | 2734 |
| 5 years                  |                 | 7752 | 7752 | 2762 | 2762 |
| 6 years                  |                 | 5326 | 5300 | 1862 | 1888 |
| 7 years                  |                 | 4091 | 4075 | 1436 | 1452 |
| 8 years                  |                 | 4101 | 4136 | 1509 | 1474 |
| 9 years                  |                 | 3969 | 3907 | 1330 | 1392 |
| Verification             | 10 years or more (11) | 38980 | 37861 | 12369 | 13488 |
| Source                   |                 | 55707 | 54881 | 18725 | 19951 | 584.3 | .0001 |
| Veriﬁed                 |                 | 33932 | 33821 | 14650 | 12761 |
| Not Verified             |                 | 28327 | 27264 | 8650 | 9713 |
| Term                     | 36              | 100834 | 95812 | 29111 | 34112 | 5332.2 | .0001 |
| 60                       |                 | 17132 | 22153 | 12914 | 7894 |
| Rent                     |                 | 52406 | 54283 | 21215 | 19338 | 497.4 | .0001 |
| Own                      |                 | 14142 | 14065 | 4934 | 5011 |
| Mortgage                 |                 | 51313 | 49515 | 15941 | 17639 |
| Other                    |                 | 105 | 103 | 35 | 37 |
| Purpose                  | Debt Consolidation | 67450 | 68676.3 | 25692 | 24465.7 | 287.6 | .0001 |
| Credit Card              |                 | 25176 | 24617.2 | 8211 | 8769.8 |
| Home Improvement         |                 | 7325 | 7070.2 | 2264 | 2518.8 |
| Major Purchase           |                 | 2455 | 2360.9 | 747 | 841 |
| Medical                  |                 | 1686 | 1659.7 | 565 | 591.3 |
| Vehicle                  |                 | 1272 | 1184.2 | 334 | 421.8 |
| Vacation                 |                 | 1109 | 1061 | 330 | 378 |
| Small Business           |                 | 1038 | 1148 | 519 | 409 |
| Moving                   |                 | 960 | 969.6 | 355 | 345.4 |
| Other                    |                 | 9495 | 9218.8 | 3008 | 3284.2 |

Table 6: Resulting analysis of statistically important continuous features in Low Credit Population.

| Feature                | Repaid Mean | Repaid St.Dev | Default Mean | Default St.Dev | T-Value | P-Value |
|------------------------|-------------|---------------|--------------|----------------|---------|---------|
| Loan Amount            | 11889.92    | 7878.93       | 14201.13     | 8403.86        | 50.73   | 0.0001  |
| Annual Income          | 75002.61    | 71771.02      | 71886.98     | 71428.30       | 7.65    | 0.0001  |
| Payment                | 383.97      | 252.34        | 449.27       | 265.13         | 44.94   | 0.0001  |
| Revolving Balance      | 13181.97    | 17183.09      | 13397.96     | 15620.37       | 2.26    | 0.0235  |
| Interest Rate          | 0.15        | 0.047         | 0.17         | 0.05           | 0.0             | 0.0001  |
| Delinquencies          | 0.01        | 0.11          | 0.08         | 0.09           | 4.90    | 0.0001  |
| Debt to Income         | 17.17       | 8.23          | 18.94        | 8.49           | 37.62   | 0.0001  |
| FICO                   | 662.45      | 2.50          | 662.40       | 2.49           | 3.08    | 0.0021  |
| Credit Lines           | 11.11       | 5.73          | 11.70        | 5.93           | 17.82   | 0.0001  |
| Inquiries              | 0.63        | 0.89          | 0.74         | 0.96           | 22.26   | 0.0001  |
| Public Records         | 0.42        | 0.82          | 0.41         | 0.78           | 2.68    | 0.074   |

predicting loan repayment from an applicant’s information. Specifically, the experiment will target the low credit applicant group in the data to contribute to the push for expansion of the peer lending space into the lower credit population. As a benchmark, we have the actual outcomes of the loans used in the experiment. We assume that any loan that was given had a peer lender with the intention of being repaid. Any loan that was not repaid was an incorrect prediction of repayment made by the lender, which negatively impacts the baseline. There were 838,429 loans repaid out of 1,055,054 total loans analyzed. In the low credit group (scores less than 670) there were 117,966 loans repaid out of 159,991. This 73% repayment rate serves as a baseline benchmark when isolating this group because 159,991 lenders expected repayment from the borrower.

5.1. ROC based model analysis

In the accuracy performance metric, the average achieved by the model based on the Two Class Bayes Point machine achieved was 75% across each of the train to test levels. This means that the Two Class Bayes Point Machine produced results aligned with true positives (loans that were repaid and predicted to be repaid) and true negatives (loans that were not repaid and predicted not to be repaid) accurately enough that these two groups made up 74% of the overall amount of data. That score also means that 26% of the data was classified as either a false positive or a false negative. The false negative count consistently stayed below 1.7% of the total observations tested at that level. False negatives are loans that were classified as not repaid when the loan was actually repaid. With false negatives the cost is higher for the borrower. This is because a lender using the model would not see a positive predicted outcome and would choose not to lend to this applicant if basing their decision solely on the prediction made. False positives are loans that were not repaid but were predicted to be repaid. For a lender, there is a higher cost associated with false positives because a loss of money occurs in these scenarios. The false positives could result in complete loss of capital if no payments are made by borrower at all or could be charged off after receiving most of the principle. Any scenario where a portion of the loan was left unpaid is not considered a successful repayment. In the false positive count 94% are loans that were partially paid back, and the other 6% were a complete loss of capital.
Since we are dealing with loan data the dataset is imbalanced. Most of the observations will have a repayment as an outcome because the normal tendency of a borrower is to repay debt, and default is the exception in this scenario. Based on this imbalance, accuracy is not the best metric for validation and instead the F1 score should be used. To further explain the reason why the accuracy can be misleading, let’s say that the model produced a classification output comprised of all positive predictions. This is not the case in the experiment, but if it were, the model would still accurately classify 117,966 loans out of the 159,991 without using a train to test split. This is a wild scenario that would not be used in an actual experiment, but the accuracy in this scenario would still be 73% which is only slightly less accurate than the actual testing performed. Hence the need to use other metrics in conjunction with the analysis. The ROC, which uses the false positive rate (10) and the true positive rate (11) is another output from the experiment showing the accuracy of the classifier.

\[
FPR = \frac{FP}{FP + TN} = \frac{33894}{33894 + 198} = .994 \tag{10}
\]

\[
TPR = \frac{TP}{TP + FN} = \frac{98898}{98898 + 257} = .997 \tag{11}
\]

The Two Class Bayes Point Machine achieved an AUC of 0.674 which can be visualized in Fig. 3. This implies that a low level of discrimination can be achieved at best, further showing that the accuracy, as defined using ROC, is not the best measure on this dataset. This is because our AUC is directly affected by class imbalance. Though this AUC is not extreme, previous work has found that an AUC of 60% or higher is desirable in information-scarce scenarios and an AUC of 70% or more would be the goal in information-rich scenarios (Berg et al., 2020).

5.2. PR based model analysis

Due to the imbalanced nature of the loan data, we used another method of classification performance evaluation. The precision (12) and recall (13) will be used to analyze the output of the precision-recall curve. Precision, which is the fraction of relevant positives out of the total positive classifications retrieved, is also shown in Table 7. On average the model was able to retrieve around 74.5% of relevant repayments, which is our precision score for the model. The recall of the model, which is the fraction of repayments retrieved from the dataset, can be seen in Table 7 as well. The model achieved a recall score of .997 in the experiment.

\[
Precision = \frac{TP}{TP + FP} = \frac{98898}{98898 + 33894} = .745 \tag{12}
\]

\[
Recall = \frac{TP}{TP + FN} = \frac{98898}{98898 + 257} = .997 \tag{13}
\]

This experiment is a simulation of a peer lender using the Two Class Bayes Point Machine to target low-credit applications classified as a future repayment. There are three different ways to look at how the lender would want to proceed with the classification problem. The lender may want to receive as many applications classified as repayments as possible, no matter how many actual result in default as they intend to cast a wide net. In this case, precision is of less importance because they would really only want to maximize recall. A peer lender can also look at the situation differently being okay with receiving less applications because they want to only receive low-credit applications classified as repayments which result in actual repayments in the future. This would mean precision would be much more important than recall and would be heavily relied upon. There are also peer lenders that are interested in receiving most of the low-credit applications classified as repayments while still avoiding most eventual defaults. This scenario would rely on a balance of both precision and recall maximization, which is ideal.

From the PR Curve shown in Fig. 4, we see that the dataset is heavily imbalanced, but the recall of a no skill model is in the range of [0.5, 1], which is far from the plotted curve. The shape of the curve indicates that the data is very randomization and complex, but imbalanced in class. The performance results from the experiment using the method are shown in Table 7. In addition to some of the metrics discussed earlier, our recall and precision, will be used to compute the F1 Score of the model (14). The F1 of the model:

\[
F1 = \frac{2(P + R)}{P + R} = \frac{2(745 + .997)}{.745 + .997} = .852 \tag{14}
\]

which takes the overall recall and precision and gives a score which balances the two. Our F1 of .856 is the score achieved when balancing precision and recall on our imbalanced dataset. A peer lender would use this score if they were interested in receiving most of the low-credit applications classified as repayments while still avoiding most even-

---

**Fig. 3. ROC for Low Credit Population using Two Class Bayes Point Machine.**

**Fig. 4. PR Curve for Low Credit Population using Two Class Bayes Point Machine.**

**Table 7. Experiment Results for Low Credit Borrowers using Two Class Bayes Point Machine.**

| Accuracy | Recall | Precision | F1 Score | AUC  |
|----------|--------|-----------|----------|------|
| 0.751    | 0.997  | 0.745     | 0.852    | 0.674 |
tual defaults. In this case, the higher the F1 score on the 0 to 9 scale, the better for the peer lender. In this imbalanced dataset our F1 is not measuring correct classifications, it is giving an edge by placing higher importance on false negatives and false positives.

6. Discussion

Four common machine learning algorithms were used for this comparison to serve as another benchmark performance for the method used in the experiment. The performance of the Decision Tree, Support Vector Machine, Logistic Regression, and Neural Network algorithms were analyzed using the same data and the same train to test ratio of a 50% split. These algorithms were selected due to their popularity and application usefulness to the binary classification problem at hand. The performance metrics were compiled and can be seen in Table 8, along with the performance of the Two Class Bayes Point Machine. Any areas in which an algorithm met or exceeded the performance of the Two Class Bayes Point Machine is indicated in Table 8. Many of the algorithms were close in performance for most of the metrics used. The Two Class Bayes Point Machine achieved a 0.751 level of accuracy, which was above the group average of 0.7424. Though the accuracy achieved was highest, the algorithms were all close in accuracy as the standard deviation of the accuracy metric was 0.005. The highest recall score was also achieved by the Two Class Bayes Point Machine at 0.997, while the mean for the group was .971 and the standard deviation was 0.0218. The Decision Tree beat all of the algorithms compared in precision with a 0.765 while the Two Class Bayes Point Machine scored slightly higher than the mean, which was 0.753. The Decision Tree also outperformed the group in AUC, but the Two Class Bayes Point Machine had a slight edge over the group mean of 0.663. From the comparison of algorithms, we can see that the Two Class Bayes Point Machine outperformed the group in 3 of the 5 metrics. Though the algorithm did not have the best scores in precision and AUC, the Two Class Bayes Point Machine did beat the group mean in both categories, taking up second place in both categories. For these reasons, we recommend the use of the Two Class Bayes Point Machine in binary classification problems, specifically in predicting repayment in peer-to-peer loans with borrowers having credit scores lower than 670.

In the feature selection process, we used t-test for continuous variables and chi squared test for the categorical variables to compare the mean of the repair and default samples. The features that were found to have statistical significance were included in the experiment, while the statistically insignificant variables were removed. The description of the continuous features that were left after the removal can be found in Table 9. These variables were also commonly found in other work as strong predictors. A comparison of the statistical significance found among the same variables in the group having credit scores of 670 or higher can be seen in the data in Table 10, which shows that the strength in these variables is not just specific to predictions in the low credit group. There have been others who have found statistical significance in many of the same variables when analyzing repayment data. In a paper that discussed the determinants of default, features were scored in the same manner to show the statistical importance to the outcome (Serrano-Cinca et al., 2015). The authors performed a t-test on each variable independently to determine impact on the loan outcome. The data classification experiment was performed with a logistic regression utilizing only the set of statistically significant variables to predict repayment. Since the t-test results showed consistency with the findings in our work this reference was included to support our reasoning in feature selection. In another paper, an analysis of defaults during economic disruption has shown that there are telling features in peer-to-peer loan data that can be used to predict their outcome (Anh et al., 2021). Both referenced papers used a logistic regression classification method with a similar subset of variables. These variables were selected based on the results of T-tests and Chi Squared evaluations. In Table 10, we will refer to the statistical significance findings of the features in both papers by “Reference A” and “Reference B”, respectively. For comparison we take the results from “Reference A” and “Reference B” to compare their statistical significance outcomes for the same variables using the sample size, the T-Value and P-Value for each continuous variable used. We also include the statistical significance found in the variables among the group with credit scores above 670 to give more coverage in showing that these variables are not just specific to our work. The consistency in the findings of statistical significance of variables across the groups in Table 10 show that the data used is coming from a commonly shared dataset, and that there is support for use of the variables selected in experiments focused on repayment predictions. These references share a

### Table 8. Performance Comparison Using Other Machine Learning Algorithms.

| Method                       | Group       | Accuracy | Recall | Precision | F1 Score | AUC   |
|------------------------------|-------------|----------|--------|-----------|----------|-------|
| Two Class Bayes Point Machine| ≥670        | 0.751    | 0.997  | 0.745     | 0.852    | 0.674 |
|                              | <670        | 0.758    | 0.978  | 0.831     | 0.854    | 0.653 |
| Decision Tree                | ≥670        | 0.741    | 0.936  | 0.765*    | 0.842    | 0.676*|
|                              | <670        | 0.808    | 0.982  | 0.817     | 0.892    | 0.716 |
| Support Vector Machine       | ≥670        | 0.739    | 0.991  | 0.741     | 0.848    | 0.627 |
|                              | <670        | 0.806    | 0.995  | 0.809     | 0.892    | 0.643 |
| Logistic Regression          | ≥670        | 0.744    | 0.968  | 0.754*    | 0.848    | 0.673 |
|                              | <670        | 0.806    | 0.980  | 0.817     | 0.891    | 0.702 |
| Neural Network               | ≥670        | 0.737    | 0.963  | 0.751     | 0.844    | 0.666 |
|                              | <670        | 0.758    | 0.878  | 0.831     | 0.854    | 0.631 |

* Met or exceeded the performance of the Two Class Bayes Point Machine on the Low Credit Group
common widely available dataset based on peer-to-peer loans from a prominent platform in the USA. This is a limitation to the study since it is only applicable to a country that utilizes the same attributes and FICO based credit scoring system. There are cases where repayment predictions in peer-to-peer lending can be performed on data from geographic regions with different credit scoring methods, but the significant variables found and used here apply to the FICO based USA model.

7. Conclusion

In this work, we have given examples of how decision-making in the peer-to-peer loan market can be aided with a machine learning approach. We have tested the viability and usefulness in repayment predictions by analyzing results and providing evidence of the potential benefits that can come out of using a data classification algorithm, the Two Class Bayes Point Machine, to build a model for prediction. We found that, when applied to historical loan repayment records, this model did not prove to be much more accurate than the baseline in prediction of the positive class (true positives vs false positives) due to the imbalance. In the repayment prediction circumstance, less focus must be given to true negatives, which has a large contribution to the accuracy metric. True negatives, though useful, have less of a cost than the false negatives and false positives when predicting loan repayment. In this case, a different metric, the F1 score must be used due to the uneven class distribution.

After adjusting for the imbalance of class, the focus on precision and recall showed that the model performed well. A precision of .745 implies that out of all of loans that were predicted to be repaid, the model correctly identified about 74.5% repayments. The cost of false positives is high in the peer lending market, but we must also look at the rate of correctly identified repayments out of the total amount of repayments that occurred because the cost of false negatives is also high in repayment prediction. The model achieved a .997 recall which shows its coverage of actual repayments. This means that the model was successfully able to recommend slightly over 99% of the applications that resulted in actual repayments. A lender would want to balance the two metrics to have confidence in using the model. A balance between the two metrics was found in the F1 score of the model which was .852 which is a positive result as this gives another measure of the model’s effectiveness.

As shown in the Table 1, a FICO score below 670 would be scrutinized by traditional lenders and denied by some that are more risk averse. When lending decisions are made utilizing the FICO score as one of the main factors, these application outcomes would result in either a denial or approval with stipulations (like a high interest rate). After being denied approval from or completely skipping traditional lenders, borrowers with fair to very bad credit scores would be the ones exploring alternate routes such as peer-to-peer lending. Only 16% of the loans approved had borrowers with a score less than 670. In the analysis 73% of all loans having an applicant with a credit score below 670 resulted in a successful repayment. Based on the experiment performed, the usage of the Two Class Bayes Point Machine by a peer lender on low credit applications can increase their chances of selecting successful applicants while avoiding those that will default.

We stated that an objective of the study was to examine the portion of the total peer-to-peer loan population that were low credit borrowers. This segment was 16% of the total loan population. This level of lending to lower credit borrowers could be the result of uncertainty due to the lack of reliable outcome prediction capabilities. The Two Class Bayes Point Machine proved to be effective in predicting loan repayment by achieving an F1 score of .852 on the low credit population of borrowers. The model, based on the Two Class Bayes Point Machine proved to be a suitable option for overcoming this uncertainty. This work supports the recommendation that the Two Class Bayes Point Machine is an option for individual peer lenders to utilize for predicting loan repayment among potential peer borrowers who have credit scores lower than 670.

Declarations

Author contribution statement

David Maloney, Sungchul Hong, PhD, Barin Nag, PhD: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data associated with this study has been deposited at https://www.kaggle.com/etbon0426/lending-club-20072020q1.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Abdallah, N.B., Destercke, S., Sallak, M., 2017. Easy and optimal queries to reduce set uncertainty. Eur. J. Oper. Res. 266 (2), 592-604.
Abdelmola, A.K., 2017. Using a naive Bayesian classifier methodology for loan risk assessment: evidence from a Tunisian commercial bank. J. Econ. Finance Adm. Sci. 22 (42), 3-24.
Abid, L., Zaghdene, S., Masmoudi, A., Ghorbel, S.Z., 2017. Bayesian network modeling: a case study of credit scoring analysis of consumer loans default payment. Asian Econ. Financial Rev. 7 (9), 846–857.

Anh, N.T., Hanh, P.T., Thu, V.T., 2021. Default in the us peer-to-peer market with Covid-19 pandemic update: an empirical analysis from lending club platform. Int. J. Entrepreneurship 25 (7), 1–19.

Banks, W., Abed, P., 1991. An efficient optimal solution algorithm for the classification problem. Decis. Sci. 22 (5), 1008–1023.

Berg, T., Burg, V., Gobovic, A., Puri, M., 2020. On the rise of fintech: credit scoring using digital footprints. Rev. Financ. Stud. 33 (7), 2845–2897.

Bhattacharya, A., Wilson, S.P., Soyer, R., 2019. A Bayesian approach to modeling mortgage default and prepayment. Eur. J. Oper. Res. 274 (3), 1112–1124.

Boiko Ferreira, L., Barddal, J., Gomes, H., Enembreck, F., 2017. Improving credit risk prediction in online peer-to-peer (P2P) lending using imbalanced learning techniques. In: Proceedings of the 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). 2017-November, Boston, MA, USA, pp. 6–9.

Chen, Y., Leu, J., Huang, S., Wang, J., Takada, J., 2021. Predicting default risk on peer-to-peer lending imbalanced datasets. IEEE Access 9, 73103–73109.

Corston-Oliver, S., Aue, A., Duh, K., Ringger, E., 2006. Multilingual dependency parsing using Bayes point machines. In: Proceedings of the Human Language Technology Conference of the NAACL Main Conference. Association for Computational Linguistics (ACL). New York City, USA, pp. 160–167. Retrieved from: https://aclanthology.org/N06-1021.

Fitzpatrick, T., Mues, C., 2021. How can lenders prosper? Comparing machine learning approaches to identify profitable peer-to-peer loan investments. Eur. J. Oper. Res. 294 (2), 711–722.

Gadzinski, G., Castello, A., 2020. Fast and frugal heuristics augmented: when machine learning quantifies Bayesian uncertainty. J. Behav. Exp. Finance 26 (100293).

Goodwin, P., Onskal, D., Stekler, H.O., 2018. What if you are not Bayesian? The consequences for decisions involving risk. Eur. J. Oper. Res. 266 (1), 238–246.

Guo, Y., Zhou, W., Luo, C., Liu, C., Xiong, H., 2016. Instance-based credit risk assessment for investment decisions in P2P lending. Eur. J. Oper. Res. 249 (2), 417–426.

Herbrich, R., Graepel, T., Campbell, C., 1999. Bayes point machines estimating the Bayes point in kernel space. In: Proceedings of IJCAI Workshop Support Vector Machines. Stockholm, Sweden, pp. 23–27.

Herbrich, R., Graepel, T., Campbell, C., 2001. Bayes point machines. J. Mach. Learn. Res. 1, 245–279.

Junier, L.M., Nardini, F.M., Renso, C., Trani, R., Macedo, J.A., 2020. A novel approach to define the local region of dynamic selection techniques in imbalanced credit scoring problems. Expert Syst. Appl. 152, 113351.

Kim, J., Cho, S., 2019. Towards repayment prediction in peer-to-peer social lending using deep learning. Mathematics 7 (11).

Kozodoi, N., Jacob, J., Lesmann, S., 2022. Fairness in credit scoring: assessment, implementation and profit implications. Eur. J. Oper. Res., 1083–1094.

Kriebel, J., Sitz, L., 2022. Credit default prediction from user-generated text in peer-to-peer lending using deep learning. Eur. J. Oper. Res. 302 (1), 309–323.

Li, A., Pericchi, L., Wang, K., 2020. Objective Bayesian inference in probit models with intrinsic priors using variational approximations. Entropy 22 (5), 513.

Liu, F., Hua, Z., Lim, A., 2015. Identifying future defaulters: a hierarchical Bayesian method. Eur. J. Oper. Res. 241 (1), 202–211.

Liu, Z., Shang, J., Wu, S.-y, Chen, P.-y., 2020. Social collateral, soft information and online peer-to-peer lending: a theoretical model. Eur. J. Oper. Res. 281 (2), 428–438.

Serrano-Cinca, C., Gutiérrez-Nieto, B., López Palacios, L., 2015. Determinants of default in P2P lending. PLoS ONE 10 (10), 22.

Short, M.B., Mohler, G.O., 2022. A fully Bayesian tracking algorithm for mitigating disparate prediction misclassifications. Int. J. Forecast., in press.

Zanin, L., 2020. Combining multiple probability predictions in the presence of class imbalance to discriminate between potential bad and good borrowers in the peer-to-peer lending market. J. Behav. Exp. Finance 100272.