Proceedings of the 4th International Conference on Computing and Data Science (CONF-CDS 2022)
DOI: 10.54254/2755-2721/2/20220667

Modelling recidivism with binary models and machine learning

Jialong Feng[1, 5,†], Lecheng Ding[2,6, †], Shuyi Li[3,7, †], Yanbo Wang[4, 8,†]

[1]School of Government, Peking University, Beijing, 100091, China
[2]Qingdao No.58 High School, Shandong Province, Qingdao, 266041, China
[3]School of mathematics and statistic, Qufu Normal University, Qufu, 273100, China
[4]Department of Econometrics and Business Statistics, Monash University, Melbourne, 3800, Australia

[5]fengjialong1999@126.com,
[6]780488185@qq.com
[7]425660803@qq.com
[8]ywan0320@student.monash.edu
[†]These authors contributed equally to this work and should be considered co-first authors.

Abstract. The article, reviewing previous study on recidivism prediction model, modeling the binary outcome of recidivism with linear regression model, probit regression model and logistic regression model. The application of machine learning techniques support vector machine (SVM) aids to compare the prediction result of each model and the relationships between recidivism probability and several personal characteristics have been identified. The prediction performances of models are similar with logistic regression model being the most parsimonious one and SVM method is tested to be not significantly superior in terms of prediction compared with traditional statistical models.

Keywords: recidivism prediction, marginal effect, SVM, regression.

1. Introduction

The criminal rate has always been a widely discussed topic all over the world since modern times and crime prediction has been recognized as an important part of the theoretical system of criminology and an indispensable prerequisite for crime prevention. The main function of crime prediction has developed from special prevention to advanced research on crime phenomena and the implementation of social crime prevention strategy. At present, countries all over the world attach importance to crime prediction work to varying degrees, and some countries have set up special crime prediction institutions.

Recidivism prediction problem has been studied for a lengthy period and historical literature has been accumulated, where increasing variables and more complicated models are introduced to better the prediction result. Not only does the recidivism prediction have practical implication for policy maker and judicial system to eliminate criminal rate and efficiently budgeting finances as it may guide former

DOI: 10.54254/2755-2721/2/20220667

prisoners who require specific consideration and supervision, but also it may provide evidence for sentencing discrimination research and gender inequality issue.

In the article, a significantly important part of crime analysis that relates to recidivism is explored: Firstly, the question on recidivism is classified as a classification problem and 6172 pieces of data from Broward County, Florida are provided to help with analysis, in which we review several previous researches on the same or similar topic and derive some useful ideas from them and then go through the data visualization. After that, A few of models including linear regression model, logistics model (logit), probit model, support vector machine model (SVM) are selected and scrutinized to help find the relationship between the probability of recidivism and the variables included and simultaneously providing better prediction results. We conclude that there has been significant relationship between recidivism and gender, age, prior convictions, misdemeanor. We identify age and female has negative effect upon the probability of recidivism whereas individuals with higher prior convictions or having committed at least one misdemeanor are more likely to reoffend. The racial difference is large between African American, Asian and other races yet the result is significantly, which is partly attributed to endogeneity. Using the classification methods and evaluation of machine learning, we then apply confusion matrix to test and compare between models. The prediction result of each model is similar and logistic model is the most appropriate and parsimonious model. Finally, we considering our limitations suggest a few improvements aspects for further research.

The remainder of this article are as followed. Section 2 traces back relevant researches especially focusing on comparison between recidivism prediction models. Section 3 exhibits the preliminary data processing procedure and result in which we discover several rudimentary relationships between variables. Section 4 outlays the basic mathematical theory upon which we constructed certain models. Section 4 explains the procedure of model construction and evaluation based on confusion matrix, Akaike information criterion (AIC) and pseudo R square. In this portion, support vector machine is introduced for complementation and as an advanced algorithms to be juxtaposed with linear model, logistics model and probability model. Python (version 3.8.6), Stata (version 17), R Studio (version 4.1.3) and other software are applied for analysis.

2. Literature review

When we analyse classification problems for justice, crime and reoffence related matters many researchers tend to use logistic model for determining whether an ex-convict is trying to commit another crime or merely having the potential of recidivism. Tollenaar and van der Heijden [1] used the data from the Dutch offender's index for the purpose of comparing the performance of eleven designated models from econometric or statistical models and machine learning models on predicting various types of recidivism based on their performance criteria. Especially, for both general recidivism and violent recidivism, the logistic model is the best candidate of predicting this type of recidivism for their dataset, as Tollenaar and van der Heijden mentioned, the logistic model does not require to strictly follow the assumption of heteroskedasticity. Their findings asserted the reality that the performance of models under machine learning's approaches does not have significant strengths over the traditional statistical models such as logistic regression and Linear Discriminant Analysis [1]. Gottfredson [2] illustrates two crime-control strategies that incurred many criticisms but increasingly demanded which are collective incapacitation and selective incapacitation both of which heavily rely on the predictive robustness and accuracy. The former one tends to impose analogous penalties or sanctions to groups of people who have convictions of homogenous offenses, providing with the future expectations of offending behavior amongst those groups. The latter one implies the sentencing guidelines that are derived from the predictive probabilities of recidivism in which they are generated by some instruments for predictions. Gottfredson also mentioned the potential bias and fairness issues that may be raised from the selection of various statistical approaches for different combinations of independent variables or for the classification result of offending. Meanwhile, the author raised the question about whether we should make use of those prediction methods at all or whether it is essential to predict for crime justice decision making. Particularly, it is very controversial in terms of crime justice decisions for individuals which might be the matter of their

DOI: 10.54254/2755-2721/2/20220667

freedom or common policies that would lead to individual decisions. Gottfredson also summarized some of the techniques that could enhance the predictive accuracy and their usefulness.

3. Data

3.1. Definition

The definition of recidivism was originally mentioned in the article written by [3]. Since our data source is a simplified version of their dataset which was modified by Adebayo [4]. Larson et al. [3] used the definition of recidivism that was provided Northpointe Inc. on their practitioner's guide as the latest crime of felony or misdemeanor recorded in the COMPAS administration date within two years. Moreover, Larson et al. exclude some of the circumstances referred as recidivism in which people who were arrested for failing to appear at their hearings in the courts, people who had infringement of traffic with tickets, or people has violations of municipal regulation or people who with a crime committed prior to the screening by the COMPAS but were arrested afterward.

3.2. Data source

We use the simplified version of recidivism data in Broward County, Florida, originally provided by ProPublica [5], and it was used Larson's article [3]. Then it was refined by Adebayo [4]. In the dataset of recidivism, we did not adjust heavily on the structure of the data. We choose variables related to actuarial recidivism result, gender, age, races, prior conviction number, whether having committed misdemeanor and other personal characteristics. Note that we did not make use of the variable "score_factor" into our models. Since this variable "score_factor" was a set of classification results about the risk of recidivism generated by a multicollinear-adjusted logistic regression model created by Adebayo as an approximated model for assessing the performance of the COMPAS model designed by Northpointe Inc. Now the company which designed the COMPAS is united by another company and renamed as 'equivant'[6]. Thus, we deleted this variable "score_factor" due to redundancy.

Here are the explanations of all variables.

Recid is a dummy variable which is equal to 1 if individual was caught committing another crime within two years of current offense, otherwise the value is 0.

Female is a dummy variable which is equal to 1 if the gender of a person is female, otherwise the value is 0.

Age is a categorical variable which is equal to 0 (labeled 'young') if the individual is under 25 at time of offense, to 1 (labeled 'middle') if the individual is between 25 and 45 at time of offense, to 2 (labeled 'old') if the individual is over 45 at time of offense.

Priors is a discrete variable which represent the number of prior convictions.

Misde is a dummy variable which is equal to 1 if current offense is a misdemeanor, otherwise the value is 0.

Race is a categorical variable which is equal to 0 if the individual is Caucasian, to 1 if the individual is African American, to 2 if the individual is Hispanic, to 3 if the individual is Asian, to 4if the individual is native American, to 5 if the individual is from other races.

3.3. Data pre-process

The method we have adopted is to firstly visualize the data provided. We use R software (R version 4.1.2 - 2021-11-01) as our visualizing tool to draw intuitive graphs. Therefore, we pre-processed the data in several steps"

Add another column of data for people aged 25 to 45. We separate the data of age into three groups as 'older than 45', 'between 25 and 45' and 'younger than 25', turning age into several dummy variables. This is because the previous data only provided two columns of data which represented people who are below 25 or who are over 45, which is not direct to obtain some main features.

DOI: 10.54254/2755-2721/2/20220667

Separate the main database into recidivism data and non-recidivism data respectively. This will help us to generate initial ideas on how to select models.

Separate races, age distributions, and gender. This helps us to classify variables.

3.4. Data pre-analysis

We firstly Summarize the main features of priors. As the only discrete random variable, the variable of prior criminals is used to assess previous criminal level ranging from 0 to 38.

Then, the data are analyzed by separating into three different columns – recidivism, non-recidivism, total with respect to different categories including age, gender, and races

Here are the tables and graphs for dummy variables.

In total, we have 6172 data, including 2809 recidivisms and 3363 non-recidivisms. Most data follow the pattern in which the number of recidivisms is generally smaller than non-recidivisms in each category because the data contain more information on non-recidivisms than recidivisms (which can be shown by graphs). However, there are some exceptions for people aged below 25 or people who belong to African America (according to graphs and tables). The probability of recidivism is especially higher in those two categories. Therefore, we may assume that age range and race can be significantly crucial to our analysis. In further analysis, we may find a more specific relationship of those variables.

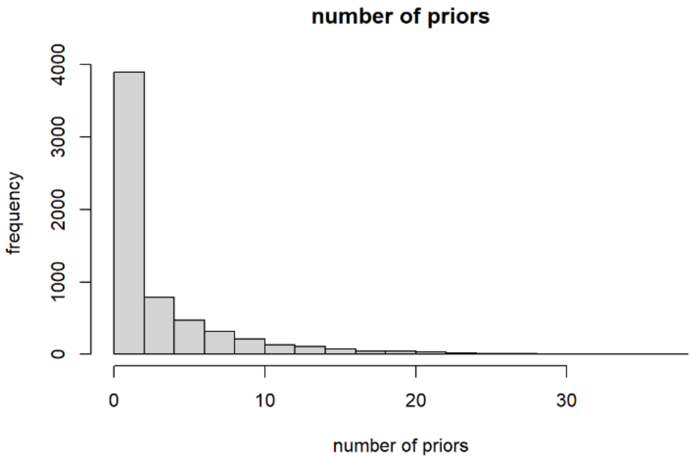

Figure 1. Histogram on "Number of Priors".

Table 1. Summary of data on recidivism, non-recidivism, and the total.

Category	Recidivism	Non-recidivism	Total
Age below 25	754	593	1347
Age between 25 and 45	1641	1891	3532
Age above 45	414	879	1293
Male	2396	2601	4997
Female	413	762	1175
Af_am	1661	1514	3175
Caucasian	822	1281	2103
Hisp	189	320	509
Asian	8	23	31
Nat_am	5	6	11
Other	124	219	343
Misdemeanour	825	1377	2202

DOI: 10.54254/2755-2721/2/20220667

3.5. Linear probability model

The statistical / econometric models we used in this article for the classification and prediction purpose of identifying the probability of recidivism by ex-convicts in the dataset are the probabilistic model and the logistic model. Therefore, the review of the theories for the probabilistic and the logistic models is essential. We will start briefly reviewing from the most basic model, which is applied for the dependent binary outcome variables, then followed by reviewing the fundamentals of the probabilistic and the logistic models.

Most of the time, the linear probability model (LPM) is often implemented in economic research for estimating binary outcome variables [7]. Specifically, it estimates the binary variable of the interest lying into one of two circumstances with a response probability. That is, suppose a variable y can have two statuses assigned with either value 1 or 0. Each state is corresponded in terms of a given probability p or 1-p for an individual i we are referring to. Formally, we can present this in a functional manner. That is,

$$\begin{cases} y_i = 1 & p \\ y_i = 0 & 1-p \end{cases} \text{for i = 1, 2, 3, …, n.} \tag{1}$$

It is desirable that we are eager to have something to include which could provide some information on explaining the movement of the variable y. Therefore, the variable y is regressed on a bunch of independent variables $x_1, x_2, x_3, \ldots, x_k$ (or regressors, these two terms can be interchangeably used). Here, we use the lower-case letter k to denote the number of regressors we incorporated in this vector form linear regression.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + e \tag{2}$$

β_0 is the unknown intercept of this regression. $\beta_1, \beta_2, \beta_3, \ldots, \beta_k$ are the unknown coefficients attached to the regressors $x_1, x_2, \ldots, x_3$ respectively. e denotes the error we have within a regression to reflect those unobserved noises which cannot be specified in the construction of a model. Therefore, we should instead replace those unknown parameters $\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_k$ by their estimates $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \ldots, \widehat{\beta_k}$, and write them as

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2 + \widehat{\beta_3} x_3 + \cdots + \widehat{\beta_k} x_k \tag{3}$$

Since the binary outcome variable y can only take discrete values 1 or 0, not any number between them. So that we instead compute their respective response probability in which it relates to either one of these two events / status, conditional on the independent variables we observed and included. Specifically,

$$P(y = 1|\boldsymbol{x}) = E(y = 1|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k \tag{4}$$

In the opposite, $P(y = 0|\boldsymbol{x}) = 1 - P(y = 1|\boldsymbol{x})$. $P(y = 1|\boldsymbol{x})$ is the response probability of the variable y in which y takes a status of interest. By default, we denote $P(y = 1\ |\boldsymbol{x})$ as the probability of "success. Here, the linear regression we specified is in vector form, and we assume that the error e satisfies the zero conditional mean property. That is, $E(e) = E(e|\boldsymbol{x}) = 0$, where $\boldsymbol{x} = x_1, x_2, x_3, \ldots, x_k$. As a result, we could estimate those unknown parameters using the OLS estimation method [7].

But Woodridge also asserted some of the limitations that the linear probability model will behave by construction. At first, the errors are not independently and identically normally distributed. Instead, they are distributed in terms of Bernoulli Distribution. So that some of the assumptions based on the Gauss-Markov theorem are not satisfied anymore. It leads to some relevant problems and as a result, applying the OLS estimation to the LPM is no longer BLUE (i.e., the OLS estimator for the parameters in the LPM is not the minimum variance (most efficient) unbiased

DOI: 10.54254/2755-2721/2/20220667

linear estimator among all the other estimators). Secondly, the interval of the response probability, $P(y=1|\boldsymbol{x})$ or $P(y=0|\boldsymbol{x})$ is not guaranteed to be lied between zero and one. In particular, this happens if we have some combinations of certain values of on these regressors. In addition to that, the "linear" property of the LPM made some of the interpretation of marginal effects of certain regressors / independent variables deviated from the real world, as every unit of increment has identical effect to the dependent variable, ceteris paribus. This does not reflect any reasonable situations of the realization. Hence, we use the probabilistic model and the logistic model as the supplements against the LPM for solving the issue of non-unity probability problem raised in the LPM.

4. Model construction

4.1. Linear regression model

Changes in socioeconomic phenomena are often affected by multiple factors, so multiple regression analysis is generally used, we refer to regression including two or more independent variables as multiple linear regression.

Considering the data of this study except priors are discrete 0-1 variables, therefore, for regression model, First by using SPSS software to observe the significance of different variables, then we use stepwise method to further improve the model, and calculate model AIC value, then introduced interaction terms, using stepwise regression, backward regression, forward regression method, and calculate the AIC value, observe the model effect, in the end we compare the four AIC values, comprehensively select the optimal model.

Which according to the regression result, the preliminary regression results that many variables are not significant, such as af_am, asian, hisp, nat_am and other, so the stepwise regression method is used to eliminate the variables. The next one is stepwise regression results.

It can be seen from the table1 that the stepwise regression is divided into six models. The observed VIF values, which is the column that is marked in red, are less than 10, so it can be considered that there is basically no collinearity, and all variables are significant. Then, we need to do outlier judgement.

4.1.1. The outlier judgement. In terms of outlier, result shows that absolute value of the deleted student residue is below 3 and cool distance is below 0.05, which implies there is no outliers. The model selection and self-correlation judgment involves the comparison of Akaike information criterion (AIC) between models forementioned.

SPSS software arranges the R square of the six models from small to large, and the model at the bottom is the optimal model, and the DW value is very close to 2. The model does not have self-correlation.

SPSS gives the residue of the model as 1322.213, and the AIC value of the stepwise regression method was 44370.55 using R software.

Considering that there may be interaction between different variables, interaction terms are introduced to this model and conducted in stepwise regression, forward and backward regression, the steps are the same as the regression model that interaction terms are not introduced.

The following are the stepwise (interaction term) regression results.

The VIF values are less than 10, and all the variables are significant.

From the result of model selection, The absolute value of Stu. Deleted residuals is below3, and the cook distance is below 0.05, so there are no outliers.

SPSS gives a residual of 1290.304 and calculates its AIC value of 44227.77 by using R software. Then, it is the forward method.

DOI: 10.54254/2755-2721/2/20220667

4.1.2. Forward method

Table 2. Forward (interaction terms).

10	（constant）	.312	.012		26.388	.000					
	priors	.061	.003	.583	20.736	.000	.291	.255	.243	.173	5.767
	young	.151	.023	.125	6.614	.000	.111	.084	.077	.380	2.629
	priorssq	-.002	.000	-.291	-10.566	.000	.208	-.133	-.124	.181	5.539
	old	-.130	.015	-.106	-8.672	.000	-.139	-.110	-.101	.913	1.095
	young * female	-.140	.035	-.055	-3.961	.000	-.032	-.050	-.046	.707	1.415
	young * af_am	.112	.026	.076	4.313	.000	.118	.055	.050	.446	2.244
	female * misde	-.041	.025	-.022	-1.626	.104	-.112	-.021	-.019	.720	1.389
	misde * asian	-.314	.133	-.028	-2.363	.018	-.040	-.030	-.028	.993	1.007
	female * af_am	-.056	.023	-.032	-2.425	.015	-.054	-.031	-.028	.773	1.293
	misde	-.028	.014	-.027	-2.091	.037	-.120	-.027	-.024	.799	1.251

The VIF, DW value given by SPSS are in the acceptable range, no collinearity and self-correlation, its residual is 1290.605 and by using R software its AIC is calculated which is 44229.21. The last one is the backward method.

4.1.3. Backward method

Table 3. Backward (interaction terms).

32	（constant）	.314	.012		26.482	.000					
	priors	.061	.003	.583	20.798	.000	.291	.256	.243	.174	5.746
	old	-.133	.015	-.109	-8.851	.000	-.139	-.112	-.103	.907	1.103
	young	.149	.023	.124	6.516	.000	.111	.083	.076	.379	2.639
	nat_am	-.330	.173	-.028	-1.910	.056	.000	-.024	-.022	.636	1.572
	misde	-.037	.012	-.035	-2.934	.003	-.120	-.037	-.034	.946	1.057
	old * asian	.260	.148	.022	1.757	.079	-.008	.022	.021	.872	1.146
	young * af_am	.113	.026	.076	4.361	.000	.118	.055	.051	.445	2.248
	young * nat_am	.856	.367	.031	2.328	.020	.020	.030	.027	.775	1.290
	young * female	-.140	.035	-.055	-3.971	.000	-.032	-.051	-.046	.714	1.401
	female * af_am	-.070	.022	-.040	-3.115	.002	-.054	-.040	-.036	.828	1.208
	female * nat_am	.663	.367	.024	1.807	.071	.020	.023	.021	.777	1.286
	female * other	-.123	.061	-.024	-2.017	.044	-.052	-.026	-.024	.976	1.024
	misde * asian	-.393	.141	-.035	-2.781	.005	-.040	-.035	-.033	.875	1.143
	priorssq	-.002	.000	-.291	-10.587	.000	.208	-.134	-.124	.181	5.525

The VIF and DW value given by the backward model, are also in the acceptable range, the steps are the same as above, it shows no self-correlation and collinearity, the residual is 1288.123, and its AIC is 44225.33 calculated by R software.

DOI: 10.54254/2755-2721/2/20220667

4.1.4. Summary for linear model

AIC price:

LPM :44370.55>forward(interaction terms):44229.21>stepwise(interaction terms):44227.77 >backward(interaction term):44225.33

After AIC comparison, the AIC of the introduced interaction term is the smallest, and the AIC of the backward method is very close to that of the stepwise method. Although the AIC of the backward method is slightly smaller than that of the stepwise method, the disadvantages of the backward method are obvious. Once a variable is eliminated from the model, it will not be introduced into the model again, and the variables calculated by the stepwise method are all significant, but some variables in the backward method are not significant. Therefore, for the regression model, The stepwise model with interaction term is effective, but its AIC value is still much larger than the probit and logit models we will discuss later, so we decide not to use this model.

4.2. Probit regression model

4.2.1. Variable selection

$$\Phi\left(\begin{array}{c} \beta_0 + \beta_1 priors + \beta_2 Old + \beta_3 Young + \beta_4 African\ American + \beta_5 Asian \\ +\beta_6 Native\ American + \beta_7 Other\ Race + \beta_8 Female + \beta_9 Misdemeanor \\ +\beta_{10} Priors^2 + \beta_{11} Priors \times Young + \beta_{12} Priors \times Other \\ +\beta_{13} Priors \times Female + \beta_{14} Old \times Asian + \beta_{15} Old \times Native\ American + \\ +\beta_{16} Young \times African\ American + \beta_{17} Young \times Native\ American + \\ +\beta_{18} Young \times Female + \beta_{19} African\ American \times Female \\ +\beta_{20} Asian \times Misdemeanor + \beta_{21} Other\ Race \times Female \\ +\beta_{22} Female \times Misdemeanor \end{array}\right) \tag{5}$$

The probit model we employ in this paper is chosen by the stepwise regression function of minimizing the Akaike information criterion. We use R square to conduct the sequential procedure of model selection. By starting with all possible and reasonable combinations of different form of independent variables, including one squared term (priors^2) and several interaction terms. Furthermore, we realized one interaction term that may cause the computation of response probability for recidivism being 0 or 1. So that we exclude that circumstance by dropping that term (Native American × Female). At the end, we end up with the model which has the following result provided in the table 1. However, there are some regressors especially in the interaction terms are extremely insignificant even at 10% significance level. Therefore, we tried to drop those highly insignificant regressors (Asian × Misdemeanor, Old × Native American and Young × Native American) sequentially from “Probit model adj-2 to “Probit model adj-4”. It turns out that the McFadden’s Pseudo R^2 continuously decreasing as we dropped those terms, and the Akaike information criterion (AIC) increases correspondingly. Noticeably, some standard errors of interaction terms become tremendously large, which is an undesirable and deteriorating situation. For instance, the AIC raised from 7495.288 to 7503.867 and the McFadden’s R^2across “Probit model adj-1” to “Probit model adj-4”, even if the Bayesian information criterion (BIC) decreased respectively as well. Usually, in econometric literatures, the use of AIC is always favored to the BIC, as it is more commonly used [8], and the AIC tends to be more tolerated in terms of diversification in econometric models than the BIC. Ultimately, employing the” model adj-1” is the best option so far, although we consider the possibility of occurring curse of dimensionality and complex computation.

4.2.2. Marginal effect of probit model. Based on the properties of marginal effect for both probit and logit model, the computation of the marginal effect of a discrete and a continuous variable for the response probability of the dependent variable varies a lot in terms of the calculation methods. But in our data, there is no existence of continuous variable. So that we can use the marginal effect of discrete variable for the response probability of recidivism only. Providing the median characteristics of each variable and controlling for other variables when we are evaluating a specific change of any characteristics, a native American tends to have 2.439% less chance of being identified as a potential recidivist than a Caucasian for every extra increase of priors. Instead, the chance of being identified as having the potential of committing another crime with increasing priors do not vary much between African American and Caucasian groups with only about 0.07353% differences. Asian group tends to be less identified as recidivists by the model with only 0.4332% lower than the that of Caucasian for every increment of priors. In terms of other unpresented races, except Hispanic, the chance of being identified as recidivist is 1.269% smaller than Caucasian when the number of priors increase by one unit. For other non-race factors, the age determinants for recidivism overall get more prominent than the gender factor and misdemeanor. The impact of gender ranks at the second place. But having misdemeanor is not a very deterministic factor for judging the probability of reoffending as number of priors increased. Those results revealed a fact that the distinction of number of priors for each individuals regardless of other characteristics may not have important implications on identifying potential recidivists.

In more general perspective, we also computed the authentic marginal effect of each factor that we incorporated in the model to the probability of recidivism at median characteristics. But the results turned out to be quite resemble to the previous marginal effect analysis of priors. Nonetheless, the age factor becomes more prominent in deciding whether those ex-convicts have the possibility of having reoffence as the age of being young tends to be identified strongly as recidivists than other age groups, holding all else constant. In this scenario, our finding under the probit model shows that the marginal effects of those race factors tend to be not quite distinctive from the age factors and gender.

Table 4. McFadden's Pseudo R^2 against the AIC and BIC.

	Model adj-1	Model adj-2	Model adj-3	Model adj-4
McFadden's Pseudo R^2	0.1242741	0.1233962	0.1230865	0.1225602
AIC	7495.2878	7500.7554	7501.3900	7503.8667
BIC	7650.01	7648.7665	7642.6733	7638.4222

Table 5. Impact of number of priors on the marginal effect of changing characteristics from its median to the response probability of recidivism.

One unit increment in priors		
Dummy variables	= 0	= 1
African American	0.04769523	0.04843051
Asian	*	0.04336305
Native American	*	0.02330555
Other Race	*	0.06038978
Old	*	0.04050328
Young	*	0.06198798
Female	*	0.05552961
Misdemeanor	*	0.04666212

(* Indicates the marginal effect of median characteristics of a recidivist on the response probability)

DOI: 10.54254/2755-2721/2/20220667

Table 6. Marginal effect of changing characteristics on the probability of recidivism at its median characteristics.

	P(recidivism = 1)	
Dummy variables	= 0	= 1
African American	0.3598889	*0.3873005
Asian	*0.3873005	0.2912190
Native American	*0.3873005	0.0942548
Other Race	*0.3873005	0.3802854
Old	*0.3873005	0.2490703
Young	*0.3873005	0.6421967
Female	*0.3873005	0.3016411
Misdemeanor	*0.3873005	0.3576093

(* Indicates the median characteristics of a recidivist)

Table 7. Probit model adj-1 (Dropping Native American*female).

	Parameter estimate:	Standard error	P-value	Confidence Interval (95%)
Intercept	-0.50914	0.04271	$< 2e^{-16}$	[-0.5929, -0.4681]
Priors	0.15410	0.00965	$< 2e^{-16}$	[0.1352, 0.1730]
Old	-0.39106	0.04523	$< 2e^{-16}$	[-0.4797, -0.3024]
Young	0.35998	0.07315	0.00786	[0.2166, 0.5034]
African American	0.07239	0.04487	0.10669	[-0.0156, 0.1603]
Asian	-0.26347	0.38333	0.49189	[-1.0148, 0.4879]
Native American	-1.02864	0.61827	0.09616	[-2.2404, 0.1832]
Other race	-0.06024	0.09789	0.53834	[-0.2521, 0.1316]
Female	0.04054	0.08538	0.63491	[-0.1268, 0.2079]
Misdemeanor	-0.07849	0.04017	0.05071	[-0.1572, 0.0002]
$Priors^2$	-0.00372	0.00045	$< 2e^{-16}$	[-0.0046, -0.0028]
Priors × Young	0.04366	0.02318	0.05961	[-0.0018, 0.0890]
Priors × Other Race	0.04187	0.02822	0.13787	[-0.0134, 0.0972]
Priors × Female	0.02241	0.01359	0.09907	[-0.0042, 0.0490]
Old × Asian	1.05635	0.62129	0.08908	[-0.1614, 2.2741]
Old × Native American	5.05951	103.02334	0.96083	[-196.87, 206.99]
Young × African American	0.24706	0.08306	0.00293	[0.0843, 0.4099]
Young × Native American	5.97415	103.03654	0.95376	[-195.98, 207.93]
Young × Female	-0.30292	0.10868	0.00531	[-0.5159, -0.0899]
African American × Female	-0.29628	0.09185	0.00126	[-0.4763, -0.1163]
Asian × Misdemeanor	-4.52393	40.68457	0.91146	[-84.27, 75.22]
Other Race × Female	-0.39072	0.22300	0.07976	[-0.8278, 0.0464]
Female × Misdemeanor	-0.14705	0.09195	0.10978	[-0.3273, 0.0332]

DOI: 10.54254/2755-2721/2/20220667

4.3. Logistic model

4.3.1. Variable selection

$$\ln\frac{\Pr(recidivism)}{\Pr(not\ recidivism)} = \beta_0 + \beta_1 priors + \beta_2 priors^2 + \beta_3 female + \beta_4 misde + \boldsymbol{\gamma}\cdot RACE + \boldsymbol{\delta}\cdot AGE + \boldsymbol{\sigma}\cdot(AGE \times female) + \varepsilon \tag{6}$$

To form a logistic regression, scholars usually select variables with significant univariate correlation with the dependent variable and contain them in model simultaneously. By a few regression operations, we sifted several variables, which is believed to be sufficiently correlated to the probability of recidivism and contribute to the accuracy of prediction. The model constructed includes basic variables collected in dataset and we add two other variables for better prediction: the square of priors and the interaction term between AGE and female. We suppose there is a non-linear relationship between the dependent variable and the prior conviction numbers and by treating it as a discrete variable, we have confirmed out thought. The result of treating *priors* variable discretely and adding quadratic term is similar so we adjust the logistic model as above for simplification.

4.3.2. Results of logistic model. The logistic regression results are shown in table. The prior conviction has a significantly positive relationship with the probability of recidivism yet the quadratic term is significantly negatively correlated with the dependent variable, which corresponds to the trend in figure which illustrates the marginal predicting probability ranging from no previous conviction to 28: As the number of previous convictions increasing, the probability of recidivism is closer to 1 yet this increasing pace is slowed down. The fluctuation at the end when conviction number is over 20 could possibly be attributed to the misspecification of model caused by lack of data for the distribution of prior conviction has a long leg. Result also shows that the accuracy of logistic model increased after adding quadratic term.

Gender is correlated with the probability of recidivism indicating man has an overtly more probable of committing a crime. However, as the result is not significant and further information for control variables are excluded in the model, it requires more valid argumentation and results to convince that there is gender discrimination or gap in terms of criminology or sentencing.

Race discrimination is a widely studied field criminologically and the sentencing data are frequently used for evidence. Economists distinguish 'taste-based' discrimination and statistical discrimination and we are focus on the latter [9]. Although evidence shows that variables are equally valid from recidivism prediction and has similar relationship with probability of recidivism between male and female, some article points out that gender itself leads to different probability of recidivism. For instance, gender-neutral (or as some remark 'male-specific') criminogenic needs are not as equally relevant as assumed when predicting recidivism [10]. In our model, we though notice some trends for example African American has remarkably higher probability of recidivism in comparison with his Caucasian counter-parts and Asian are least likely to reoffend among 6 races group holding other factors fixed but these results are not statistically significant. The predictive margins of different are also different according to figure yet with a large difference of confidence interval due to disproportionately racial data, which is doubted that differences of probability are actually caused by discriminatory sentencing or higher violence rate.

Table 8. Regression result of logistic regression.

recid	Coefficient	Std. err.	z	P>z	[95% conf.	interval]
priors	0.268333	0.01465	18.32	0	0.239619	0.297047
priors_sq	-0.00649	0.000717	-9.06	0	-0.00789	-0.00508
female	-0.11901	0.174735	-0.68	0.496	-0.46148	0.223463
misde	-0.18211	0.059406	-3.07	0.002	-0.29854	-0.06567
race						

DOI: 10.54254/2755-2721/2/20220667

Table 8. (continued).

African American	0.07889	0.063363	1.25	0.213	-0.0453	0.203078
Hispanic	-0.15426	0.110046	-1.4	0.161	-0.36994	0.06143
Asian	-0.54369	0.433249	-1.25	0.21	-1.39284	0.305463
native American	-0.26066	0.642886	-0.41	0.685	-1.5207	0.999372
others	-0.11924	0.129113	-0.92	0.356	-0.3723	0.133814
age						
middle	-0.92725	0.078103	-11.87	0	-1.08033	-0.77417
old	-1.571	0.100481	-15.63	0	-1.76793	-1.37406
female#young	-0.68706	0.228918	-3	0.003	-1.13573	-0.23839
female#middle	-0.06856	0.197363	-0.35	0.728	-0.45538	0.318267

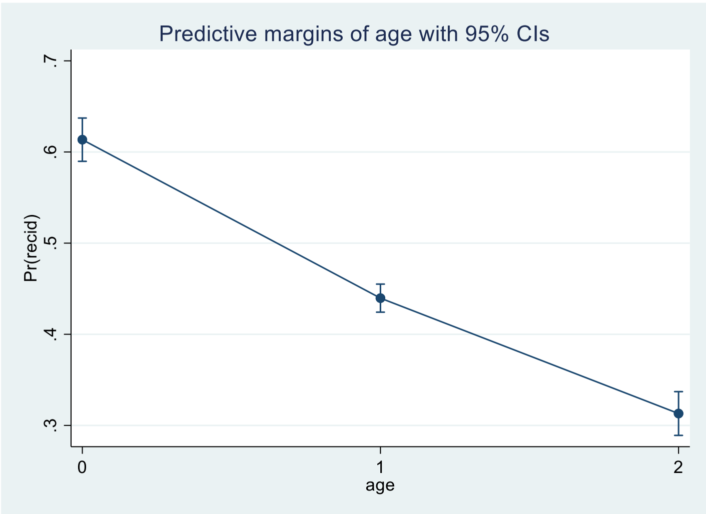

Figure 2. Predictive margins of age with 95% CIS.

5. Evaluation

5.1. Comparison between models

We based several information criteria and measurement of performance to assist us selecting the desirable model specification that might be appropriate for our dataset as well as for measuring predictability of recidivism when we implement or fit the model into different sample or different data. Typically, we would use mainly adopt the Akaike information criteria (AIC) and the hit and miss table for analyzing the balance between the degree of parsimony and goodness-of-fit as well as the general predictability of our models.

The initial and the most intuitive way of looking how the goodness of fit of various models that we have specified is by using the Pseudo R^2. Veall and Zimmermann [11] reasserted the fact that when it comes to the diagnostic tests for models in the context of either limited dependent variable or continuous dependent variable, the R^2 methods should be avoided to use. One of the possible reasons of that advice is that the underlying mathematical structure of various R^2 methods require the computation of sum of the squared residuals (SSR) in which they are mostly derived by the OLS estimation method. However, in the nonlinear model context such as the probit and logit models, the sum of squared residuals is not able to be computed, and they typically use the maximum likelihood estimation method. Therefore, as the alternative measures of goodness-of-fit, various types of Pseudo-R^2 approaches are increasingly popular to be used in the diagnostic context when it comes to the case of limited dependent variable models, including the models where its dependent variable is discrete such as the probit and logit models. Generally, a Pseudo-R^2 is a goodness-of-fit measure applied to the limited dependent variable or the discrete dependent variable case, which is resemble to the standard R^2-approach in the linear regression or the continuous data scenario [12]. Among these academically proposed methods of Pseudo-R^2, we preliminarily interested in adopting the McKelvey and Zavoina's Pseudo R^2 which was proposed by McKelvey and Zavoina [13]. Since if we evaluate the model performance across the linear probability model in which it used the OLS estimation method, the logit and probit models in which they both used the maximum estimation method with the binary dependent variable (i.e., having recidivism or not), the McKelvey and Zavoina's Pseudo R^2 should reflect the best comparability as a good approximation to the standard R^2 which is based on the OLS estimation method under the latent variable framework [12]. However, Hagle and Mitchell [14] in his findings demonstrated that the results of the McKelvey and Zavoina's Pseudo R^2 and other R^2 measures may not necessarily hold when the model misspecification occurs (i.e., when the errors are no longer normally distributed such as skewed or binomial), compared to the case of normally distributed error term. Veall and Zimmermann [11] stated that in that situation, McFadden's R^2 might be a better R^2 measure to be adopted as it is less subject to variation. Our histogram of residuals generated by the probit and logit models illustrated binodal distribution respectively.

Table 9. Robust Jarque-Bera Test for the probit and the logit model.

Model type:	Probit model	Logit model
Jarque-Bera test statistic:	457.17	459.71
Degrees of freedom:	2	2
P-value:	$< 2.2e^{-16}$	$< 2.2e^{-16}$

We could also observe that both robust Jarque-Bera tests for probit and logit models reject the null hypothesis that the errors are normally distributed at 5% significance level. These two results together with the sampling distributions of the residuals justify the fact that the models are misspecified, and they do not comply with the assumption. Thus, the use of the McFadden's R^2 as the measure of goodness-of-fit could be proposed. To be detailed, the McFadden's R^2s reported for the probit and logit models are respectively 0.1242741 and 0.1186815. This implies that the probit model fits the data slightly better than the one fitted by the logit model. Keeping into account the fact that the McFadden's Pseudo R^2 might not be a good approximation (unbiased) to the OLS - R^2, mentioned by Veall and Zimmermann [11], the adjusted R^2 of the linear probability model that we have specified before is about 0.1556. This claims that the LPM may have better goodness-of-fit over the probit and the logit model without incorporating the correction of potential biasness existed in the Pseudo R^2. Thus, it imposes some degree of difficulties for evaluating model performance across different models with different estimation methods and different model assumptions based on R^2-type measures. However, Hagle and Mitchell [14] asserted the importance of joint use of Pseudo R^2 measure with other types of measures of model performance. We also reported the results of Akaike information criteria and Bayesian information criteria as well as the accuracy of prediction for recidivism under different measures such as the hit and miss tables.

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) both seek a balance between the goodness-of-fit and parsimony as high level of model complexity is always undesirable due to the curse of dimensionality. Maydeu-Olivares and Garcia-Forero [12] restated the properties of the AIC and BIC that they aim to penalize for extra addition of parameters. Normally, the smallest value of the AIC or the BIC should be often preferred. However, the BIC prefers more in terms of model parsimony with respect to the number of parameters than the AIC. So that the BIC has stricter penalty on adding additional parameters than the AIC. A very typical set of the AIC and BIC formulae have been provided below.

$$AIC = -2 \times LogLik + 2 \times k \tag{7}$$

$$BIC = -2 \times LogLik + k \times \ln{(n)} \tag{8}$$

Where LogLik is the log-likelihood of the estimated model, k is the number of parameters of the estimated model and n is the sample size or the number of observations in the data. The only situation where we are indifferent from looking at either the AIC or the BIC is when the number of observations equal to approximately 7.40, which is unrealistic and unreasonable to have such number and the sample size corresponding to it.

The AIC favor the probit model (7495.28775) more than the logit model (7524.86056). Because the conduct of the probit model specification is under the procedure of stepwise regression in which it aims to minimize the AIC, given the initial provision of different independent variables. Therefore, in conjunction with the higher McFadden's Pseudo R^2 relative to that of the logit model, we may consider that the probit model might be preferred in accordance with this perspective. In contrast, the BIC favors the logit model more than the probit model due to parsimonious structure (i.e., a smaller number of parameters). Exclusively, the AIC (7879.301) and BIC (7960.034) of the LPM has the largest value among these three model specifications. Kuha[8] stated that even though both AIC and BIC have their own objective and motivation of defining a sense of "good model", they usually have desirable approximations to their theoretical perspective outcomes, and this typically implies that they could specify a desirable model given the observed data of the realization. However, this condition does not always hold. That is why we are required to provide more realistic measure of the model performance than those classic measurements (e.g., the prediction accuracy table or the hit and miss table).

The prediction accuracy table is provided below. To better our prediction, we additionally used support vector machine (SVM) as machine learning algorithms in recidivism prediction for robustness test. In machine learning, support-vector machines (SVMs, also support-vector network) are supervised learning category models with associated learning algorithms which analyze data for classification and regression analysis. Developed at AT&T Bell Laboratories by Vladimir Vapnik and his colleagues SVMs has been broadly applied in predicting marketing future state [15], credit default [16], regional economic outcome [17] and many other social and economic domains. SVM is one of the most robust prediction methods and able to converge at a fast rate [18], being based on statistical learning frameworks or VC theory proposed by Vapnik and Chervonenkis [15]. SVM learning is superior in handling the entangled complex relationship between predictors and target variable especially in absence of theory. Despite that there have been numerous articles trying to introduce machine learning method in criminology, less researchers apply SVM learning machines in recidivism prediction. We compare the result of this advanced method with traditional statistical model to test the validation and performance of latter. The application of SVM does not bring apparently superior performance to traditional logit or linear model measured against confusion matrix which confirm the validation and effectiveness of conventional algorithms as for recidivism prediction.

We apply python for SVM application coding and randomly separate the dataset into two parts: One with 5000 individuals' characteristics is training set and the rest is compiled as test set. SVM model classify by forming several hyperplanes and choose the reasonable one as the best that represents the largest separation, or margin, between the two classes. In other words, the hyperplane allows the distance from it to the nearest data point on each side to be maximized. We repeat the above steps several times and retain the means for comparison.

We could observe that there are some noticeable differences between the overall accuracy and the recall of specifying recidivism or not with different baselines from the probit and the logit model. To be specific, the logit model has overall 2.44% higher ability of correct prediction on whether the ex-convicts have recidivism or not, and 9.22% higher percentage of correctly identified those ex-convicts who were having recidivism than that of the logit model at the time of the data collection. But it has a little bit weaker performance with respect to correctly identifying ex-criminals who have no recidivism than that of the probit model, which is about 3.81% distinction. In particular, the probit model has relatively robust performance with respect to correctly identifying those ex-criminals who have not reoffended with 94.17% of the correctly identified rate, compared to that value of 90.36% predicted by the logit model. However, all these two models are less capable of identifying true recidivists than correctly identifying those who had not reoffended at the time of data collection.

The LPM has the highest rate of identifying those true recidivists among other two models with about 55.46%. On the other hand, this model might not be very excellent at the identification of non-recidivists from the group of ex-convicts that do not have reoffence. Generally, the level of accuracy is not quite sensitive in the context of different model specifications. But the level of recall and precision do vary under different model specifications.

SVM model has a precision rate of 53.2% and a recall rate of 41.5%. The accuracy rate, being 63.0% is not significantly different from logit or probit model. SVM algorithms though invented and proved to pertain complicated relationship between independent variables and predictive target shows no superiority in our research.

5.2. Model selection

Even though the prediction outcome against the reality was presented most closely in the linear probability model, due to the limitations of the linear probability model that was discussed by Woodridge [7], Westin [19] also pointed out some issue that the linear probability model possessed in nature and he excluded the use of the linear probability model from his paper. This in turn has similar circumstances that we have already faced in our analysis. Westin [19] also reported the issue that may arise from the employment of the probit model in which its probabilistic distribution would have the form of inverse cumulative normal function which would imply intractable prediction computation. Based on these studies, we have the reasonable ground to employ the logistic model to be the best candidate in our data analysis as it prevents these potential problems that the linear probability model and the probit model would suffer from. In addition to that, the logistic model is more parsimonious than the probit model and we do not observe any significant distinction of the prediction outcome presented in the hit and miss table or the confusion matrix that we discussed before under these two models. Furthermore, a parsimonious model is more likely to prevent the negative effect from the curse of dimensionality and the computation would be much easier than a complicated model such as the probit model we specified in our paper. Another well-explained intuition is that a model has perfect fitness to the data would not necessarily imply a robust predictability or forecasting ability on another set of data.

Table 10. Hit-and-Miss table for the LPM

Recidivism	Pr(Recidivism=0 |X)	Pr(Recidivism=1 |X)	Sum
= 0	2600	763	3363
= 1	1251	1558	2809
Sum	23851	2321	6172

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{2600\ +\ 1558}{6172} = \frac{4158}{6172} = 0.6737$$

$$Positive\ predictive\ rate\ (Precision) = \frac{TP}{TP + FP} = \frac{1558}{2321} = 0.6713$$

$$Negative\ predictive\ rate = \frac{TN}{TN + FN} = \frac{2600}{3851} = 0.6751$$

$$True\ positive\ rate\ (Recall)\ = \frac{TP}{TP + FN} = \frac{1558}{2809} = 0.5546$$

$$True\ negative\ rate = \frac{TN}{TN + FP} = \frac{2600}{3363} = 0.7731$$

DOI: 10.54254/2755-2721/2/20220667

Table 11. Hit-and-Miss table for the Probit model.

Recidivism	Pr(Recidivism=0 |X)	Pr(Recidivism=1 |X)	Sum
= 0	3167	196	3363
= 1	2108	701	2809
Sum	5275	897	6172

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{701 + 3167}{6172} = \frac{3868}{6172} = 0.6267$$

$$Positive\ predictive\ rate\ (Precision) = \frac{TP}{TP + FP} = \frac{701}{897} = 0.7815$$

$$Negative\ predictive\ rate = \frac{TN}{TN + FN} = \frac{3167}{5275} = 0.6004$$

$$True\ positive\ rate\ (Recall)\ = \frac{TP}{TP + FN} = \frac{701}{2809} =\ 0.2496$$

$$True\ negative\ rate = \frac{TN}{TN + FP} = \frac{3167}{3363} = 0.9417$$

Table 12. Hit-and-Miss table for the Logit model.

Recidivism	Pr(Recidivism=0 |X)	Pr(Recidivism=1 |X)	Sum
= 0	3039	324	3363
= 1	1849	960	2809
Sum	4888	1284	6172

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{960 + 3039}{6172} = \frac{3999}{6172} = 0.6511$$

$$Positive\ predictive\ rate(Precision) = \frac{TP}{TP + FP} = \frac{960}{1284} = 0.7477$$

$$Negative\ predictive\ rate = \frac{TN}{TN + FN} = \frac{3039}{4888} = 0.6217$$

$$True\ positive\ rate\ (Recall)\ = \frac{TP}{TP + FN} = \frac{960}{2809} =\ 0.3418$$

$$True\ negative\ rate = \frac{TN}{TN + FP} = \frac{3039}{3363} = 0.9036$$

6. Limitations and improvements

We try to construct a model with accuracy to predict one's probability of recidivism based on his background and quantify the effect of gender, races, age, misdemeanors and numbers of prior conviction on probability of recidivism. We achieve a level of approximately 71-72% of prediction accuracy and a true positive rate (TPR) of XXX which is a considerable result but there are still some improvements to be made for our prediction model and causal relationship identification.

As for dependent variables, the data collected in the article are mainly whether one have reoffended in two years. This may be not a perfect predictive indicator of recidivism as it has been statistically proved that the periods before individuals reoffend vary depending on the multiple level of violence of

DOI: 10.54254/2755-2721/2/20220667

their first crime, behaviors in prison and metal health condition. As a result, survival models or event history analysis in sociology are generally introduced in recidivism prediction to separate subgroups of samples (those who reoffend and who not) for creating multiple dependent variables. For instance, split population of whether one reoffends and the duration between [18]. With the inclusion of further detailed data of recidivists' duration, we can better our prediction by applying more complicated and accurate models. Another issue is able to be solved by more detailed information requires more specific classification of crime. Criminologically, recidivism data are generally processed and classified into different basic division: general recidivism, violent recidivism, sexual recidivism etc. Recent studies have identified the limitation of certain methods of prediction when apply to different offender population for multiple characteristics, which may strongly influence the consequences of recidivism, will be inappropriately standardized and separate analysis and comparison between different categorized population may better the prediction results [1,20].

The correlation coefficient of most of the races in all of the models are not significant and no valid evidence of discrimination towards minority group is drawn from our study. However, we do not infer that the judicial sentences against certain races do not exist as there are other omitted variables for example income, social status, educational level and types of crime offended. Omission of theses variables in model may lead incorrect conclusion of discrimination due to endogeneity and scholars may implement alternative methods as quasi-experimental approach [21], Difference in difference or regression discontinuity design. As the probability of reoffending is often self-selection due to personal characteristics, more advanced experiments must be designed to avoid endogeneity problems.

7. Conclusion

The study reviews previous researches over recidivism prediction ranging from classical regression models to advanced algorithm and then form three models separately for comparison of econometric interpretability and predictive performance. The relationships between selected variables relevant to offenders' physical and social characteristics and recidivism probability are similar in different models: negative result shown in female and age whereas misdemeanor has positive relationship. The effect of number of prior convictions upon reoffending possibility is more accurately depicted by adding quadratic term reflecting a reverse u-shape curve and racial difference are ambiguous due to variable omission and other possible limits.

To compare between models, a combination of pseudo-R square, AIC and BIC is the basic standard against which we define whether a model is 'good enough'. All of the three indices for each model fail to show distinct disparity with each model performing better in respective index: R square of LPM model is higher than the pseudo-R square of logit mode and probit model; logit regression has the lowest AIC and probit has the lowest BIC. None of the differences of the three indices are large enough for us to jettison any models. By applying SVM machine learning method, we complete a robustness check and confirm the validity of traditional statistical method in terms of recidivism prediction as machine learning's result are similar with few further inductions.

Acknowledgement

Jialong Feng, Lecheng Ding, Shuyi Li and Yanbo Wang contributed equally to this work and should be considered co-first authors.

References

[1] Tollenaar, N., & van der Heijden, P. G. M. (2013). Which method predicts recidi-vism best?: a comparison of statistical, machine learning and data mining predictive models. Journal of the Royal Statistical Society. Series A (Statistics in Society), 176(2), 565–584.
[2] Gottfredson, D. M. (1987). Prediction and classification in Criminal Justice Deci-sion making. Crime and Justice, 9, 1–20. https://doi.org/10.1086/449130
[3] Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). How We Ana-lyzed the COMPAS Recidivism Algorithm. ProPublica. https://www.propublica.org/article/how-we-

analyzed-the-compas-recidivism-algorithm
[4] Adebayo, J. (2017b). propublica_data_for_fariml [Unpublished raw data]. ProPub-lica. Retrieved May 14, 2022, from https://github.com/adebayoj/fairml/blob/master/doc/example_notebooks/propublica_data_for_fairml.csv
[5] ProPublica (n.d.). compass-analysis [Unpublished raw data]. ProPublica. Retrieved May 15, 2022, from https://github.com/propublica/compas-analysis
[6] Equivant (2019, April 4). Practitioner's Guide to COMPAS Core. equivant. https://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf
[7] Woodridge, J. M (2020). Introductory Econometrics: A Modern Approach (7th ed.). Cengage Learning.
[8] Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. Sociological Methods & Research, 33(2), 188–229. https://doi.org/10.1177/0049124103262065
[9] Lang, K., & Spitzer, A. K.-L. (2020). Race Discrimination: An Economic Perspec-tive. The Journal of Economic Perspectives, 34(2), 68–89.
[10] van der Knaap, L. M., Alberda, D. L., Oosterveld, P., & Born, M. Ph. (2012). The Predictive Validity of Criminogenic Needs for Male and Female Offenders: Comparing the Relative Impact of Needs in Predicting Recidivism. Law and Human Behavior, 36(5), 413–422.
[11] Veall, M. R., & Zimmermann, K. F. (1996). PSEUDO-R^2 MEASURES FOR SOME COMMON LIMITED DEPENDENT VARIABLE MODELS. Journal of Economic Surveys, 10(3), 241–259. https://doi.org/10.1111/j.1467-6419.1996.tb00013.x
[12] Maydeu-Olivares, A., & García-Forero, C. (2010). Goodness-of-Fit Testing. Inter-national Encyclopedia of Education, 190–196. https://doi.org/10.1016/b978-0-08-044894-7.01333-6
[13] McKelvey, R. D., & Zavoina, W. (1975). A Statistical Model for the Analysis of Ordinal Level Dependent Variables. The Journal of Mathematical Sociology, 4(1), 103–120. https://doi.org/10.1080/0022250x.1975.9989847
[14] Hagle, T. M., & Mitchell II, G. E. (1992). Goodness-of-Fit Measures for Probit and Logit. American Journal of Political Science, 36(3), 762–784. https://doi.org/10.2307/2111590
[15] Cui, D., & Curry, D. (2005). Prediction in Marketing Using the Support Vector Machine. Marketing Science, 24(4), 595–615.
[16] Moula, F. E., Guotai, C., & Abedin, M. Z. (2017). Credit default prediction model-ing: an application of support vector machine. Risk Management, 19(2), 158–187.
[17] Hu, Z. (2020).Economic Forecast of Coastal Regions Based on Support Vector Machine Model. Journal of Coastal Research, 211-214.
[18] Blanchard, G., Bousquet, O., & Massart, P. (2008). Statistical Performance of Support Vector Machines. The Annals of Statistics, 36(2), 489–531.
[19] Westin, R. B. (1974). PREDICTIONS FROM BINARY CHOICE MODELS. Jour-nal of Econometrics, 2(1), 1–16. https://doi.org/10.1016/0304-4076(74)90027-x
[20] Glover, A. J., Nicholson, D. E., Hemmati, T., Bernfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high-risk feder-al offenders. Criminal justice and behavior, 29(3), 235-249.
[21] Everett, R. S., & Wojtkiewicz, R. A. (2002). Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing. Journal of Quantitative Criminology, 18(2), 189–211.