

互联网金融背景下客户认购定期存款预测分析

张小宝

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2024年8月16日; 录用日期: 2024年9月9日; 发布日期: 2024年11月19日

摘要

本文在互联网金融蓬勃发展的背景下, 针对客户认购定期存款的行为预测进行了深入研究。鉴于互联网金融环境下客户行为数据的复杂性和不确定性, 本文创新性地提出了一种基于密度权重与直觉模糊孪生支持向量机的鲁棒分类器模型。该模型通过引入密度权重来优化样本空间的分布, 增强了对不均衡数据集的处理能力; 同时, 结合直觉模糊集理论, 有效处理了数据中的噪声和异常值, 提升了模型的预测精度。在互联网金融竞争加剧的当下, 准确预测客户认购定期存款的行为, 有助于银行及金融机构优化产品设计、制定精准的营销策略, 进而提升客户满意度和市场份额。实验结果表明, 该分类器显著提高了客户认购定期存款预测的准确性和鲁棒性, 为银行精准识别和服务客户群体提供了有力工具。

关键词

互联网金融, 客户认购定期存款, 密度权重, 支持向量机

Analysis of the Forecast of Customers' Subscription to Time Deposits in the Context of Internet Finance

Xiaobao Zhang

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Aug. 16th, 2024; accepted: Sep. 9th, 2024; published: Nov. 19th, 2024

Abstract

In this paper, an in-depth study is conducted on the prediction of customers' behaviour in subscribing time deposits in the context of the booming development of Internet finance. In view of the complexity and uncertainty of customer behavioural data in the Internet financial environment, this paper innovatively proposes a robust classifier model based on density weights and intuitionistic

fuzzy twin support vector machine. The model optimises the distribution of the sample space by introducing density weights, which enhances the ability to handle unbalanced datasets; at the same time, combining with the intuitionistic fuzzy set theory, it effectively handles the noise and outliers in the data, and improves the prediction accuracy of the model. At a time of intensified competition in Internet finance, accurately predicting the behaviour of customers subscribing to time deposits helps banks and financial institutions to optimize product design and formulate precise marketing strategies, which in turn enhances customer satisfaction and market share. The experimental results show that the classifier significantly improves the accuracy and robustness of the prediction of customer subscription to time deposits, and provides a powerful tool for banks to accurately identify and serve their customer groups.

Keywords

Internet Finance, Customers' Subscription to Time Deposits, Density Weights, Support Vector Machines

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在全球经济持续繁荣与金融市场日益深化的浪潮中，个人与企业的财富管理需求呈现出前所未有的多元化趋势，这驱动着银行业不断追求创新，以灵活应对市场的风云变幻。银行定期存款，这一历史悠久的金融瑰宝，凭借其无可比拟的安全性与稳定的收益表现，始终在金融市场中稳坐重要席位，深受客户信赖。然而，中国经济发展步入新常态，金融监管体系日臻完善，行业内的竞争亦愈发激烈，加之互联网金融的异军突起，共同编织了一幅复杂多变的商业银行经营图景[1]。面对如此挑战，各大银行纷纷使出浑身解数，通过提升服务品质、打造创新金融产品、优化客户体验等多维度策略，以期在激烈的市场竞争中脱颖而出。在定期存款领域，银行更是下足了功夫。它们不仅通过提高存款利率的竞争来吸引客户眼球，还巧妙设计了灵活多变的存款期限、简化了繁琐的办理流程，并推出了个性化的服务方案，力求全方位满足客户的多样化需求。与此同时，金融科技的蓬勃发展为定期存款市场注入了前所未有的活力，使得客户能够享受到更加便捷、高效且智能化的金融服务体验。值得注意的是，银行客户认购定期存款的行为并非孤立事件，而是深受经济金融环境变迁、市场竞争格局演变、客户需求多样化以及政策环境支持等多方面因素的共同影响。特别是在全球经济不确定性加剧的当下，投资者的风险偏好普遍趋于保守，对资金的安全性和稳定性需求愈发强烈，这无疑为银行定期存款市场的持续增长提供了强大的动力。此外，互联网金融的迅猛发展以及各种理财产品的层出不穷，也在悄然改变着银行客户的存款认购行为。这种变化要求银行必须更加深入地理解客户需求，不断优化产品设计，制定精准有效的营销策略，以提升客户满意度和忠诚度。同时，建立科学的预测模型，以准确把握客户是否愿意认购定期存款的动向，对于银行而言同样至关重要。这不仅有助于银行更好地规划资源配置，还能在激烈的市场竞争中抢占先机。

支持向量机(Support Vector Machines, SVM)是由 Vapnik 和 Cortes 提出的一种以统计学习理论和最优化理论为基础构建的机器学习方法[2]。随着 SVM 理论与算法日趋成熟，其广泛引用于分类、回归预测、图像识别等领域[3]-[5]。然而传统的 SVM 往往遭受计算负担，为了克服这一困难，Suresh [6]等人提出孪生支持向量机(Twin SVM, TSVM)，与 SVM 不同的是，TSVM 通过在正负数据类别中分别寻找超平面来

减轻计算负担,从而使 TSVM 的运行速度比传统 SVM 快四倍。进一步, S. Rezvani [7]等人提出了一种新的直觉模糊孪生支持向量机(IFTSVM), IFTSVM 基于 IFN 为每个训练样本分配一对隶属度和非隶属度函数,这两个度量减少噪声对模型的影响,并从噪声中识别出支持向量,从而提高分类精度。由于传统的分类算法往往假设数据集平衡,当数据集不平衡时,传统的分类器得到的分离超平面往往偏向于多数类,使得少数类分类准确率降低。在实际应用中,数据集大多是不平衡的。在这些不平衡的数据集中,往往更关注少数类的分类情况。在数据集不平衡的情况下,主要有两种流行的方法来提高分类性能:抽样方法和加权方法。采样方法涉及两类,欠采样方法和过采样方法[8] [9]。第二种方法是对训练数据点赋予合适的权重,相关工作见[10] [11]。另外由于支持向量数据描述(SVDD)其灵活地描述边界且不需要对数据分布进行假设而被广泛使用, Cha 等人[12]提出了一种称为密度加权(DW)支持向量数据描述(DWSVDD)的新模型。在 DWSVDD 中,权重生成算法受到密度概念的启发,引入了密度权重的概念,密度权重是使用 k 近邻方法,根据目标数据的密度分布得出的每个数据点的相对密度。受 IFTSVM 和密度权重的启发,我们提出了一种新的分类器(DWIFTSVM)来解决二分类问题。

受 DW, IFTSVM 的启发,本文提出了一种新的基于密度权重的模糊孪生支持向量机(DW-IFTSVM)模型应用于对客户是否认购定期存款的预测模型,提高预测精度。改进的基于基密度权重的模糊孪生支持向量机对异常值和不平衡数据具有鲁棒性,为银行更加精确地识别客户是否认购定期存款提供了方案。

2. 预备知识

在本章中,对 SVM、CDFTSVM、FSVM 的理论进行简要的介绍,为了方便,我们假设 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 是训练集,样本矩阵 $X \in R^{n \times p}$ 是由有 p 个指标的 n 个观测值组成,其中 $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T \in R^p$ 表示第 i 个训练样本, $y_i = \{-1, +1\}$ 表示 x_i 所属类别。正类样本矩阵 $X_+ \in R^{n_+ \times p}$ 以及负类样本矩阵 $X_- \in R^{n_- \times p}$, 其中 n_+ 和 n_- 分别代表正类样本数和负类样本数。

2.1. 软间隔支持向量机

线性情况下,给定训练样本 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, SVM 的分类超平面为: $\omega^T x + b = 0$, 其中法向量 $\omega \in R^p$, $b \in R$ 为偏置项。SVM 优化模型为:

$$\begin{aligned} \min_{\omega, b, \xi} & \frac{1}{2} \|\omega\|_2^2 + C e^T \xi \\ \text{s.t.} & \begin{cases} D(X\omega + eb) - e + \xi \geq 0, \\ \xi \geq 0. \end{cases} \end{aligned} \tag{1}$$

其中 $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ 是松弛向量, C 是非负的惩罚参数, $D = \text{diag}(y_1, y_2, \dots, y_n)$, e 和 0 分别是元素全为 1 和 0 的 n 维列向量。根据拉格朗日乘子法以及 KKT 条件可知(1)式的对偶问题为:

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \alpha^T D X X^T D \alpha - e^T \alpha \\ \text{s.t.} & \begin{cases} 0 \leq \alpha \leq C e, \\ e^T D \alpha = 0. \end{cases} \end{aligned} \tag{2}$$

其中 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ 是拉格朗日乘子向量。若求得(2)式的最优解 α , 即可确定超平面 $\omega^T x + b = 0$ 。若给定一个新的样本点 x_{new} , 可根据如下决策函数对其进行判别:

$$f(x) = \text{sgn}(\omega^T x_{new} + b) \tag{3}$$

在非线性情况下，定义映射 $\phi(\cdot): x \rightarrow \phi(x)$ ，通过引入核函数 $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ 求解优化问题，核函数能够将原本无法分离的数据变得可分，然后在高维空间中利用内积运算计算分类面以及样本点之间的距离，从而处理非线性分类问题。

2.2. 模糊支持向量机

噪声和异常值在许多真实世界的数据集中是普遍存在的，而标准的 SVM 对每个样本点的处理是平等的，因此对噪声和离群值比较敏感。为了解决这一问题，提出了 FSVM 算法，对不同的训练样本赋予不同的模糊隶属度，然后找到最优的分离超平面将数据分为两类。给定训练集 T ，隶属度函数定义为：

$$\mu(x_i) = \begin{cases} 1 - \frac{\|\phi(x_i) - C^+\|}{r^+ + \delta}, & y_i = +1, \\ 1 - \frac{\|\phi(x_i) - C^-\|}{r^- + \delta}, & y_i = -1. \end{cases} \quad (4)$$

其中， $\delta > 0$ 是一个可调参数，用于避免 $\mu(x_i) = 0$ ， r^+ (r^-) 和 C^+ (C^-) 代表正(负)类的半径和中心，定义如下：

$$\text{正类样本和负类样本类中心的定义分别为: } C^+ = \frac{1}{n_+} \sum_{y_i=+1} \phi(x_i), \quad C^- = \frac{1}{n_-} \sum_{y_i=-1} \phi(x_j),$$

$$\text{正类和负类的类半径定义如下: } r^+ = \max_{y_i=+1} \|\phi(x_i) - C^+\|; \quad r^- = \max_{y_i=-1} \|\phi(x_j) - C^-\|.$$

为了方便，记 $\mu(x_i) = \mu_i$ ，FSVM 表达式如下：

$$\begin{aligned} \min_{\omega, b, \xi} & \frac{1}{2} \|\omega\|_2^2 + C \sum_1^n \mu_i \xi \\ \text{s.t.} & \begin{cases} D(X\omega + eb) - e + \xi \geq \mathbf{0}, \\ \xi \geq \mathbf{0}. \end{cases} \end{aligned} \quad (5)$$

2.3. 坐标下降模糊孪生支持向量

考虑到 FSVM 和 TSVM 的优点，Gao 等人在 2015 年提出了 CDFTSVM，其表达式如下：

$$\begin{aligned} \min_{\omega_+, b_+, \xi_-} & \frac{1}{2} \|X_+ \omega_+ + e_+ b_+\|^2 + \frac{c_1}{2} \|\omega_+\|^2 + c_3 s_-^T \xi_- \\ \text{s.t.} & -(X_- \omega_+ + e_- b_+) + \xi_- \geq e_-, \quad \xi_- \geq \mathbf{0}, \end{aligned} \quad (6)$$

和

$$\begin{aligned} \min_{\omega_-, b_-, \xi_+} & \frac{1}{2} \|X_- \omega_- + e_- b_-\|^2 + \frac{c_2}{2} \|\omega_-\|^2 + c_4 s_+^T \xi_+ \\ \text{s.t.} & (X_+ \omega_- + e_+ b_-) + \xi_+ \geq e_+, \quad \xi_+ \geq \mathbf{0}. \end{aligned} \quad (7)$$

其中 $s_+ \in R^{n_+}, s_- \in R^{n_-}$ 是模糊向量，其中 $\xi_+ \in R^{n_+}, \xi_- \in R^{n_-}$ 是松弛向量。

3. 密度权重直觉模糊孪生支持向量机

为了减少噪声和离群点的影响，选择一个合适的隶属度函数对于 FSVM 以及 CDFTSVM 是十分重要的。如图 1 所示：边界上的两个点 A 和 B 对于 -1 类拥有相同的隶属度，但是他们对分类效果的贡献是不一样的，在这种情况下，FSVM 很难将噪声与支持向量区分开，因此引入直觉模糊数。

3.1. 线性情况

给定训练集 T ，非隶属度函数定义为： $\nu(x_i) = (1 - \mu(x_i))\rho(x_i)$ ，其中 $0 \leq \mu(x_i) + \nu(x_i) \leq 1$ ， $\rho(x_i)$ 有如下定义：

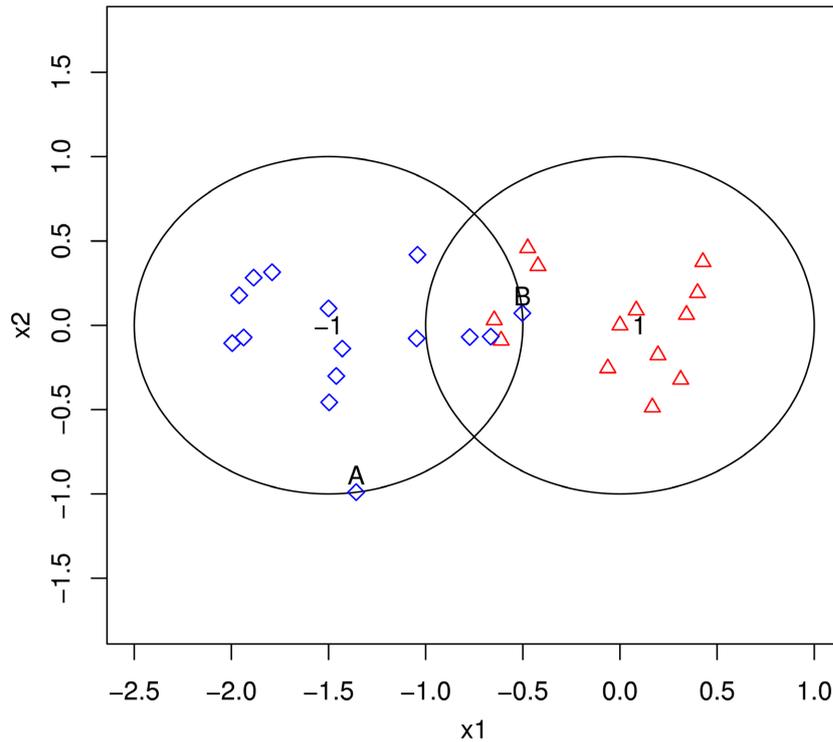


Figure 1. Similar degree of membership for two training samples.
图 1. 隶属度相同的两个训练样本

$$\rho(x_i) = \frac{|\{x_j \mid \|\phi(x_i) - \phi(x_j)\| \leq \alpha, y_i \neq y_j\}|}{|\{x_j \mid \|\phi(x_i) - \phi(x_j)\| \leq \alpha\}|}, \quad (8)$$

其中 $\alpha > 0$ 是一个可调参数， $|\cdot|$ 表示集合的势。方便起见，记 $\nu(x_i) = \nu_i$ 。

通过计算隶属度与非隶属度，可以通过下式得分函数计算样本点的得分值，

$$s_i = \begin{cases} \mu_i, & \nu_i = 0, \\ 0, & \mu_i \leq \nu_i, \\ \frac{1 - \nu_i}{2 - \mu_i - \nu_i}, & \text{others.} \end{cases} \quad (9)$$

线性情况下，DWIFTSVM 的目标函数如下：

$$\begin{aligned} \min_{\omega_+, b_+, \xi_-} & \frac{1}{2} \|X_+ \omega_+ + e_+ b_+\|^2 + \frac{c_1}{2} \|\omega_+\|^2 + c_3 s_-^T D_- \xi_- \\ \text{s.t.} & -(X_- \omega_+ + e_- b_+) + \xi_- \geq e_-, \xi_- \geq \mathbf{0}, \end{aligned} \quad (10)$$

和

$$\begin{aligned} \min_{\omega_-, b_-, \xi_+} & \frac{1}{2} \|X_- \omega_- + e_- b_-\|^2 + \frac{c_2}{2} \|\omega_-\|^2 + c_4 s_+^T D_+ \xi_+ \\ \text{s.t.} & (X_+ \omega_+ + e_+ b_+) + \xi_+ \geq e_+, \xi_+ \geq \mathbf{0}. \end{aligned} \tag{11}$$

其中 c_1, c_2, c_3, c_4 均是非负的惩罚参数, ξ_+, ξ_- 是松弛向量, $s_+ \in R^{n_+}, s_- \in R^{n_-}$ 分别是正类样本和负类样本的得分向量, $D_{\pm} = \text{diag}(d_i)$ 分别表示正类和负类样本的密度权重对角矩阵, 其中

$$d_i = D(x_i) = 1 - \frac{d(x_i, x_i^k)}{\max_{j \in \text{trainset}} d(x_j, x_i^k)}, \quad x_i^k \text{ 表示 } x_i \text{ 的 } k \text{ 个近邻点, } d(x_i, x_i^k) \text{ 表示 K-NN 距离, 且}$$

$$d(x_i, x_i^k) = \frac{1}{k} \sum_{j=1}^k d(x_i, x_i^{kj}), \quad x_i^{kj} \text{ 表示 } x_i \text{ 的 } k \text{ 个近邻点的第 } j \text{ 个。构造(10)式拉格朗日函数如下:}$$

$$\begin{aligned} L(\omega_+, b_+, \xi_-, \alpha, \beta) &= \frac{1}{2} \|X_+ \omega_+ + e_+ b_+\|^2 + \frac{c_1}{2} \|\omega_+\|^2 + c_3 s_-^T D_- \xi_- \\ &+ \alpha^T [e_- - \xi_- + (X_- \omega_- + e_- b_-)] - \beta^T \xi_-. \end{aligned} \tag{12}$$

其中 $\alpha > 0, \beta > 0$ 为拉格朗日乘子, 我们对(12)式求偏导令其为 0 有:

$$\begin{aligned} \frac{\partial L}{\partial \omega_+} &= X_+^T (X_+ \omega_+ + e_+ b_+) + c_1 \omega_+ + X_-^T \alpha = 0, \\ \frac{\partial L}{\partial b_+} &= e_+^T (X_+ \omega_+ + e_+ b_+) + e_-^T \alpha = 0, \\ \frac{\partial L}{\partial \xi_-} &= c_3 D_-^T s_- - \alpha - \beta = \mathbf{0}. \end{aligned} \tag{13}$$

由(13)式可知 $\begin{pmatrix} X_+^T \\ e_+^T \end{pmatrix} (X_+ \quad e_+) \begin{pmatrix} \omega_+ \\ b_+ \end{pmatrix} + \begin{pmatrix} X_-^T \\ e_-^T \end{pmatrix} + c_1 E_1 \omega_+ = 0$ 。令 $H_+ = [X_+, e_+]$, $H_- = [X_-, e_-]$, $u_{\pm} = \begin{pmatrix} \omega_{\pm} \\ b_{\pm} \end{pmatrix}$,

则有 $u_+ = -(H_+^T H_+ + c_1 E_1)^{-1} H_+^T \alpha$ 。故而(10)式对偶问题如下:

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \alpha^T H_- (H_+^T H_+ + c_1 E_1)^{-1} H_+^T \alpha - \alpha^T e_- \\ \text{s.t.} & \mathbf{0} \leq \alpha \leq c_3 D_-^T s_- \end{aligned} \tag{14}$$

同理可知(11)式对偶问题为:

$$\begin{aligned} \min_{\beta} & \frac{1}{2} \beta^T H_+ (H_-^T H_- + c_2 E_2)^{-1} H_-^T \beta - \beta^T e_+ \\ \text{s.t.} & \mathbf{0} \leq \beta \leq c_4 D_+^T s_+ \end{aligned} \tag{15}$$

式中 $E_1 = E_2 = \begin{bmatrix} I \\ 0 \end{bmatrix}$ 。一旦(14)~(15)式得到解决, 则可获得两个类的超平面。若给定一个新的数据点, 可根据下述决策函数对数据进行判别, 决策函数如下:

$$f(x) = \arg \min_{i=\pm 1} \left\{ \frac{|\omega_+^T x_{new} + b_+|}{\|\omega_+\|}, \frac{|\omega_-^T x_{new} + b_-|}{\|\omega_-\|} \right\}. \tag{16}$$

3.2. 非线性情况

在非线性情况下, 两个非平行超平面可以被表示为:

$$k(x, X^T) \omega_+ + b_+ = 0, \quad k(x, X^T) \omega_- + b_- = 0, \quad \text{其中 } k(x_1, x_2) = (\phi(x_1), \phi(x_2)) \text{ 为核函数, } X = (X_+^T \quad X_-^T)^T。$$

通过引入核函数，非线性的 DWIFTSVM 可以表示为：

$$\begin{aligned} \min_{\omega_+, b_+, \xi_-} & \frac{1}{2} \|k(X_+, X^T)\omega_+ + e_+ b_+\|^2 + \frac{c_1}{2} \|\omega_+\|^2 + c_3 s_-^T D_- \xi_- \\ \text{s.t.} & -(k(X_-, X^T)\omega_+ + e_+ b_+) + \xi_- \geq e_-, \xi_- \geq \mathbf{0}, \end{aligned} \tag{17}$$

和

$$\begin{aligned} \min_{\omega_-, b_-, \xi_+} & \frac{1}{2} \|k(X_-, X^T)\omega_- + e_- b_-\|^2 + \frac{c_2}{2} \|\omega_-\|^2 + c_4 s_+^T D_+ \xi_+ \\ \text{s.t.} & (k(X_+, X^T)\omega_- + e_- b_-) + \xi_+ \geq e_+, \xi_+ \geq \mathbf{0}. \end{aligned} \tag{18}$$

同理线性情况，令 $G_+ = [k(X_+, X^T) \ e_+]$ ， $G_- = [k(X_-, X^T) \ e_-]$ ， $v_{\pm} = \begin{pmatrix} \omega_{\pm} \\ b_{\pm} \end{pmatrix}$ 。且 $v_+ = -(G_+^T G_+ + c_1 E_1)^{-1} G_+^T \alpha$ ， $v_- = -(G_-^T G_- + c_2 E_2)^{-1} G_-^T \beta$ 。则(17)~(18)式对偶问题如下：

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \alpha^T G_- (G_+^T G_+ + c_1 E_1)^{-1} G_-^T \alpha - \alpha^T e_- \\ \text{s.t.} & \mathbf{0} \leq \alpha \leq c_3 D_-^T s_-, \end{aligned} \tag{19}$$

和

$$\begin{aligned} \min_{\beta} & \frac{1}{2} \beta^T G_+ (G_-^T G_- + c_2 E_2)^{-1} G_+^T \beta - \beta^T e_+ \\ \text{s.t.} & \mathbf{0} \leq \beta \leq c_4 D_+^T s_+. \end{aligned} \tag{20}$$

一旦上述问题得到解决，便可直接得到两个非并行超平面，若给定新的数据点，可由如下决策函数进行判定：

$$f(x) = \arg \min_{\pm} \frac{|k(x_{new}, X^T)\omega_{\pm}^T + b_{\pm}|}{\sqrt{\omega_{\pm}^T k(X, X^T)\omega_{\pm}}} \tag{21}$$

4. 基于 DWIFTSVM 的客户认购定期存款预测模型

4.1. 数据来源与变量描述

本文的数据来自 UCI 数据库(<https://archive.ics.uci.edu/>)，该数据集是一家葡萄牙银行机构 2008 年 5 月到 2010 年 11 月的所有话访活动市场调查结果的银行客户相关数据，用来预测银行客户是否认购其定期存款并将其分类。本文的响应变量为客户是否会认购定期存款，一共有两种可能的状态：认购和不认购。综合考虑数据结构和指标特征，在众多影响因素中主要考虑了以下 9 个解释变量对响应变量进行建模分析，这些变量均与客户是否认购定期存款有高度相关关系[13]。这些变量可分为 4 类：客户情况、与银行关系、接触银行活动状况和经济社会环境状况。具体表现为

- (1) 客户情况：年龄、婚姻状况、房贷、个人贷款。
- (2) 与银行关系：银行月利率。
- (3) 基础银行活动状况：活动期间被联系次数。
- (4) 经济社会环境状况：就业变化率、消费者信心指数、居民消费价格指数。

4.2. 定期存款预测模型的建立

在本节中进行实证研究，将提出的 DW-IFTSVM 模型应用于银行客户是否认购定期存款预测，关于

模型参数的设置, 对于 SVM, 设置 $C = \{2^{-8}, 2^{-7}, \dots, 2^7, 2^8\}$; 对于 TSVM、IFTSVM 和 DW-IFTSVM, $c_j = \{2^i \mid i = -8, -7, \dots, 7, 8\}$, $j = (1, 2, 3, 4)$, 同时令 $c_1 = c_3, c_2 = c_4$ 。在非线性情况, 考虑高斯核函数, 核参数 $\sigma \in \{2^i \mid i = -8, -7, \dots, 7, 8\}$ 。对于所有模型参数的选择, 本文使用十折交叉验证法来搜索最优参数。本文的评价指标选择准确率(ACC), 为了消除实验随机性的影响, 进行 100 次数值实验, 计算评价指标的平均值和标准差(sd)。考虑实际数据为非线性可分, 所以建立高斯核函数下非线性可分模型, 模型结果如表 1 所示。提出的模型提高了预测准确率, 将其应用于客户是否认购定期存款预测模型可以带来更好的效果。

Table 1. Data set prediction results

表 1. 数据集预测结果

	SVM	TSVM	IFTSVM	DW-IFTSVM
	Acc. \pm sd	Acc. \pm sd	Acc. \pm sd	Acc. \pm sd
数据集	0.7588 \pm 0.325	0.7322 \pm 0.426	0.7601 \pm 0.3027	0.7987 \pm 0.227

5. 结论

本文首先在引言部分对互联网背景下的客户认购定期存款以及支持向量机的现状做了一个简明扼要的描述, 其次对文章所涉及的部分理论做了简单的阐述。在第三节, 本文提出了一种新的密度加权直觉模糊孪生支持向量机模型, 旨在提升了银行客户定期存款认购预测的准确性和效率, 并对其理论做了详细的介绍。在实证部分, 采用新提出的模型与传统的分类器模型对客户认购定期存款的准确率进行比较, 结果表明: 提出的模型提高了预测准确率, 将其应用于客户是否认购定期存款预测模型可以带来更好的效果, 该模型不仅为银行精准营销和服务提供了有力支持, 也为金融领域的数据挖掘和智能决策提供了新的思路和方法。在未来研究中可进一步探索更多高级机器学习算法与直觉模糊集的融合, 以及在不同金融场景下的应用效果。

参考文献

- [1] 秦玉芳. 定期存款利率下调银行成本压力居高不下[N]. 中国经营报, 2022-05-02(B01).
- [2] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1007/bf00994018>
- [3] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [4] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [5] Chapelle, O., Haffner, P. and Vapnik, V.N. (1999) Support Vector Machines for Histogram-Based Image Classification. *IEEE Transactions on Neural Networks*, **10**, 1055-1064. <https://doi.org/10.1109/72.788646>
- [6] Jayadeva, Khemchandani, R. and Chandra, S. (2007) Twin Support Vector Machines for Pattern Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**, 905-910. <https://doi.org/10.1109/tpami.2007.1068>
- [7] Rezvani, S., Wang, X. and Pourpanah, F. (2019) Intuitionistic Fuzzy Twin Support Vector Machines. *IEEE Transactions on Fuzzy Systems*, **27**, 2140-2151. <https://doi.org/10.1109/tfuzz.2019.2893863>
- [8] Li, P., Qiao, P. and Liu, Y. (2008) A Hybrid Re-Sampling Method for SVM Learning from Imbalanced Data Sets. 2008 5th International Conference on Fuzzy Systems and Knowledge Discovery, Jinan, 18-20 October 2008, 65-69. <https://doi.org/10.1109/fskd.2008.407>
- [9] Akbani, R., Kwek, S. and Japkowicz, N. (2004) Applying Support Vector Machines to Imbalanced Datasets. *Machine Learning: ECML 2004 15th European Conference on Machine Learning*, Pisa, 20-24 September 2004, 39-50. https://doi.org/10.1007/978-3-540-30115-8_7
- [10] Huang, Y.-M. and Du, S.-X. (2005) Weighted Support Vector Machine for Classification with Uneven Training Class

- Sizes. 2005 *International Conference on Machine Learning and Cybernetics*, Vol. 7, 4365-4369. <https://doi.org/10.1109/icmlc.2005.1527706>
- [11] Ji, M. and Xing, H. (2017) Adaptive-Weighted One-Class Support Vector Machine for Outlier Detection. 2017 *29th Chinese Control and Decision Conference (CCDC)*, Chongqing, 28-30 May 2017, 1766-1771. <https://doi.org/10.1109/ccdc.2017.7978802>
- [12] Cha, M., Kim, J.S. and Baek, J. (2014) Density Weighted Support Vector Data Description. *Expert Systems with Applications*, **41**, 3343-3350. <https://doi.org/10.1016/j.eswa.2013.11.025>
- [13] 张利利, 郭淑妹, 马艳琴. 基于数据挖掘技术的银行客户定期存款认购模型研究[J]. *数学的实践与认识*, 2019, 49(21): 95-102.