Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.25153

A Compliance Service Mode Selection Model Based on Machine Learning Algorithms

Zhongkai Niu[1†], Weilin Li[1†], Zhuyin Yan[2], Huali Chen[1*]

[1]*School of Law, Guangzhou Xinhua University, Dongguan, China*
[2]*Taishan College, Shandong University, Jinan, China*
†These two authors contributed equally to this work and should be considered as co-first author.
**Corresponding Author. Email: 375921677@qq.com*

Abstract. This study is based on multidimensional compliance feature data of enterprises, and deeply explores and constructs an intelligent selection model for compliance service modes based on machine learning algorithms. By systematically comparing the performance of decision trees, random forests, support vector machines (SVM), BP neural networks, and the non-linear weight particle swarm optimization based support vector machine (NWPSO-SVM) classification model proposed in this study on the same dataset, the empirical results clearly show that the NWPSO-SVM model exhibits excellent and comprehensive classification ability in compliance service mode selection tasks. The model achieved the best results in all key evaluation metrics, with an accuracy of 0.894, a recall of 0.894, a precision of 0.886, and a stable F1 Score of 0.886. This series of significantly leading indicator values not only confirms the high accuracy of the model's prediction results. By utilizing this model for precise analysis and pattern prediction of enterprise compliance characteristics, potential compliance risk points can be identified proactively at the beginning of data service launch or product design, and appropriate service patterns can be matched. This enables compliance requirements to be more intelligently and efficiently embedded at the source of business processes, achieving a transition from passive response to active prevention, effectively improving the accuracy and effectiveness of pre regulation, reducing compliance costs and violation risks in the later stage, and laying a solid technical foundation for building a secure, reliable, and trustworthy data service ecosystem. In summary, the NWPSO-SVM model has been proven to be an ideal tool for intelligent and precise selection of enterprise compliance service models due to its significantly superior comprehensive performance.

Keywords: Compliance service mode, non-linear weighted particle swarm algorithm, support vector machine.

1. Introduction

With the rapid development of the digital economy, data resources have gradually become the core element driving social operation and business innovation. However, the large-scale collection,

DOI: 10.54254/2755-2721/2025.25153

circulation, and utilization of data also pose risks such as privacy breaches, algorithmic discrimination, and cross-border security. In order to balance data value mining and risk prevention, a legal framework centered on pre regulation has gradually been formed globally [1]. For example, the EU's General Data Protection Regulation (GDPR) establishes the principle of "Privacy by Design", requiring companies to embed data protection mechanisms during product development; The Chinese Data Security Law and Personal Information Protection Law establish prior obligations such as data classification and grading, risk assessment, and compliance review, emphasizing the standardization of data activities from the source. The current research focuses on how to deeply integrate legal rules with technological governance and explore the effectiveness boundaries of pre regulation. For example, scholars focus on the implementation path of algorithm transparency, data minimization principle, and the conflict and coordination of regulatory logic between different jurisdictions [2]. The research difficulty lies in the fact that the speed of technological iteration far exceeds the legal update cycle, making it difficult for regulatory tools to cover new data application scenarios [3]. At the same time, jurisdictional disputes over cross-border data flows have intensified the complexity of pre regulation.

Machine learning algorithms provide technical support for predicting compliance service models by automatically analyzing massive amounts of legal texts and business data. On the one hand, models based on natural language processing (NLP) can parse legal terms and contract texts, construct dynamically updated compliance knowledge graphs, and assist enterprises in identifying compliance risks in real-time in data collection, storage, sharing, and other aspects. By using entity recognition technology to automatically label sensitive personal information and combining it with a rule engine to determine whether it meets the "informed consent" requirement of the Personal Information Protection Law. On the other hand, supervised learning and unsupervised learning algorithms can predict the probability of compliance risks in specific scenarios [4]. Train classification models using historical penalty data to predict whether cross-border data transmission behavior of enterprises triggers regulatory review; Identify implicit compliance vulnerabilities through cluster analysis. In addition, reinforcement learning can simulate regulatory game environments and optimize the dynamic adjustment ability of compliance strategies [5]. This article explores a compliance service mode selection model based on machine learning algorithms using multi-dimensional compliance feature data of enterprises.

2. Data set introduction

The dataset used in this article contains multidimensional compliance feature data from 628 enterprises, aiming to study the key factors that affect the selection of data compliance service models. The dataset covers 10 feature variables and 1 predictor variable (compliance service mode: basic/advanced/customized), including enterprise size, privacy sensitivity, compliance budget, regulatory pressure index, daily data processing volume, industry risk level, game ability index, legal strictness score, past violation frequency, and other indicators. This dataset can reflect the dynamic game relationship between enterprise compliance needs, risk tolerance, and legal constraints. Select some datasets for display, as shown in Table 1.

DOI: 10.54254/2755-2721/2025.25153

Table 1. Partial dataset

Bargaining power	Company size	Compliance budget	Data volume	Industry risk	Past violations	Privacy sensitivity	Regulatory pressure	Compliance service
0.24	306.21	25.68	99.08	2	1	0.78	4	basic
0.08	360.20	26.71	30.94	2	0	0.24	5	basic
0.93	339.47	22.15	3.98	2	2	0.72	5	advanced
0.58	366.38	19.12	60.80	1	3	0.97	4	advanced
0.67	348.87	24.98	221.62	2	1	0.95	4	advanced
0.91	317.06	5.66	84.28	3	1	0.60	4	advanced

3. Method

3.1. Nonlinear Weighted Particle Swarm Optimization algorithm

Nonlinear Weighted Particle Swarm Optimization (NWPSO) is an intelligent optimization algorithm that improves upon the classical Particle Swarm Optimization (PSO) algorithm. The core idea is to balance the algorithm's ability between global exploration and local development by dynamically adjusting inertia weights, thereby improving optimization efficiency. Unlike the fixed or linearly decreasing weights of standard PSO, NWPSO adopts a non-linear weighting strategy, which enables the algorithm to maintain strong global search ability in the early stages of iteration to explore a vast solution space, and gradually enhance local fine search ability in the later stages of iteration to accurately approximate the optimal solution. This dynamic adjustment mechanism is more in line with the search requirements at different stages of complex optimization problems [6].

The design of nonlinear weights is the core innovation of the algorithm. A typical method is to generate a weight curve through mathematical functions, which shows a smooth decreasing trend as the number of iterations increases. In the initial stage, the weight is relatively high, and particles can quickly traverse the potential solution area with a large step size, avoiding getting stuck in local optima; As the iteration progresses, the weights gradually decrease according to a nonlinear law, and the speed of particles gradually slows down, thus conducting a refined search within the possible optimal solution neighborhood. This non-linear regulation is more flexible than linear variation, which can avoid premature convergence and prevent search stagnation caused by low weights in the later stage [7]. At the same time, algorithms typically retain the information guidance mechanisms of individual and group optima to ensure the effectiveness of search direction.

3.2. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm based on statistical learning theory, whose core idea is to classify or regress data by constructing a maximum interval hyperplane. For linearly separable problems, the goal of SVM is to find a hyperplane that can separate data points of different categories and maximize the distance from the hyperplane to the nearest sample point [8]. This "interval maximization" strategy not only ensures the robustness of classification, but also reduces the sensitivity of the model to noise, thereby improving generalization ability. When the data is linearly inseparable, SVM uses kernel techniques to map the raw data to a high-dimensional feature space, making it linearly separable in the high-dimensional space. Gaussian kernel function can achieve nonlinear classification by implicitly calculating the

DOI: 10.54254/2755-2721/2025.25153

inner product of high-dimensional space without explicitly defining complex mappings. In addition, SVM introduces the concept of soft interval, allowing some samples to violate interval constraints, balancing classification accuracy and model complexity by adjusting parameter C, and avoiding overfitting. The model principle of support vector machine is shown in Figure 1.

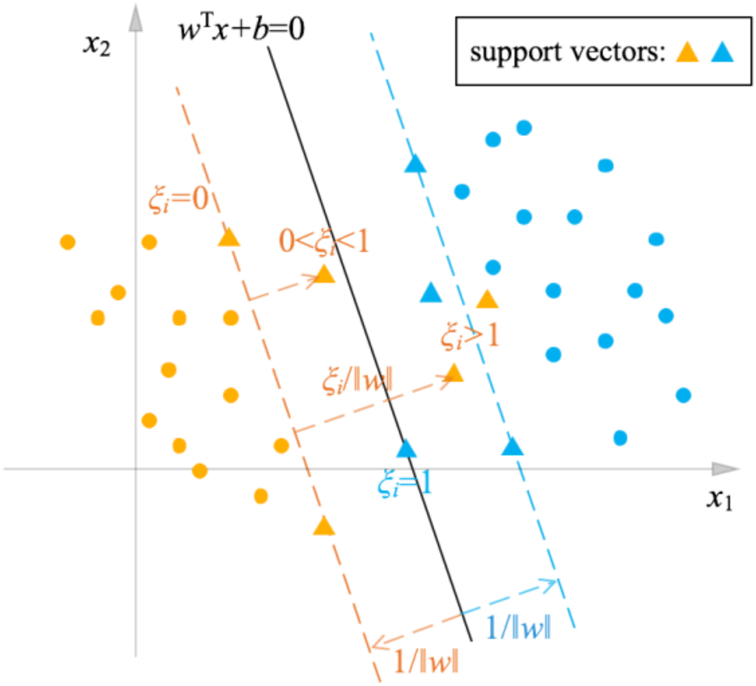

Figure 1. Support Vector Machine model

3.3. Support Vector Machine for Nonlinear Weighted Particle Swarm Optimization algorithm

The core principle of optimizing Support Vector Machines (SVM) using Nonlinear Weighted Particle Swarm Optimization (NWPSO) algorithm is to adaptively adjust the key parameters of SVM through intelligent search strategy to improve the classification performance of the model. The performance of SVM is highly dependent on parameter selection, and traditional grid search is inefficient and prone to getting stuck in local optima [9]. NWPSO uses the cross validation accuracy of SVM as the fitness function and iteratively searches the parameter space through particle swarm optimization. Each particle represents a set of parameter combinations, and its position updates are guided by individual historical optima and group optima. Nonlinear inertia weights dynamically balance global exploration and local development: in the initial stage, high weights allow the particle swarm to extensively traverse the parameter space, avoiding missing potential optimal regions; As the iteration progresses, the weights decrease nonlinearly, and the particles gradually focus on the high fitness area for fine search, thereby accurately locating the optimal parameters. This process effectively solves the problems of high computational cost and premature convergence in traditional parameter tuning, while combining the mapping ability of kernel functions to enhance the adaptability of SVM to complex data distributions [10].

4. Result

In this experiment, the non-linear weighted particle swarm optimization (PSO) algorithm was used to optimize the support vector machine (SVM) classification model. PSO was set with 50 particles, a

DOI: 10.54254/2755-2721/2025.25153

maximum of 200 iterations, and non-linear decreasing inertia weights (0.9 → 0.4). SVM was evaluated using 5-fold cross validation, and the dataset was divided into training and testing sets at a ratio of 7:3, with evaluation metrics including accuracy, precision, recall, and F1 score; The hardware configuration is Intel i7-12700H processor, 32GB memory/1TB SSD, and the software environment is Ubuntu 22.04 system, relying on Python 3.10 and scikit learn 1.3.0.

Classification experiments were conducted using decision trees, random forests, support vector machines, BP neural network models, and the proposed classification model based on nonlinear weight particle swarm optimization for support vector machines. The results are shown in the Table 2.

Table 2. Model evaluation

Model	Accuracy	Recall	Precision	F1
decision tree	0.757	0.757	0.766	0.761
random forest	0.831	0.831	0.821	0.825
SVM	0.556	0.556	0.638	0.583
BP neural network	0.815	0.815	0.664	0.732
Our model	0.894	0.894	0.886	0.886

The comparison results of various indicators are shown in Figure 2.

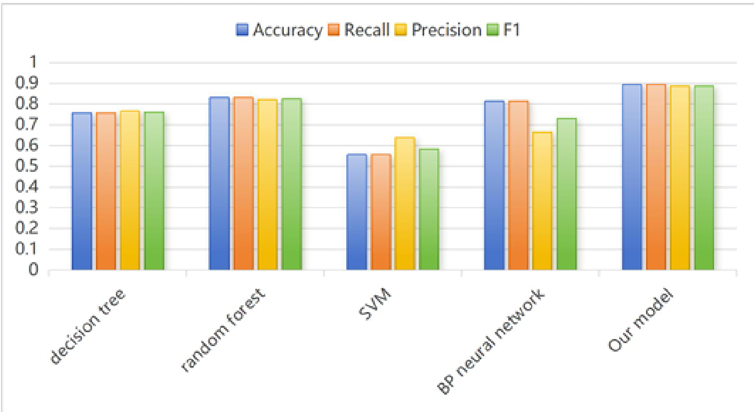

Figure 2. Model evaluation

According to the comparison of the provided classification performance indicators, Our model demonstrates significant advantages, achieving the highest values in all evaluation indicators: Accuracy (0.894), Recall (0.894), Precision (0.886), and F1 Score (0.886), indicating its best comprehensive classification ability and accurate and comprehensive prediction. In contrast, the performance of random forest and BP neural network models is second, with accuracy exceeding 0.81, but there are differences: random forest has more balanced indicators (F1=0.825), while BP neural network's Precision (0.664) is significantly lower than its Accuracy (0.815) and Recall (0.815), suggesting that there may be more false positives, resulting in a relatively low F1 Score (0.732). The decision tree model performs moderately, with various indicators ranging from 0.75 to 0.76 (F1=0.761). The SVM model has the weakest performance, with an accuracy and recall of only 0.556. Although Precision (0.638) is slightly higher, its lower recall results in a significantly lower F1 Score (0.583) compared to other models, indicating a higher missed detection rate. Overall, our

DOI: 10.54254/2755-2721/2025.25153

model's classification performance in compliance service mode selection tasks is significantly better than other compared models.

5. Conclusion

This study focuses on multidimensional compliance feature data of enterprises, and explores and constructs a compliance service mode selection model based on machine learning algorithms. To comprehensively evaluate the effectiveness of the model, we systematically compared the decision tree, random forest, support vector machine (SVM), BP neural network, and the support vector machine (NWPSO-SVM) classification model optimized based on nonlinear weight particle swarm algorithm proposed in this study. The rigorous experimental results clearly demonstrate that the proposed NWPSO-SVM model exhibits comprehensive and significant advantages in classification performance. Specifically, the model achieved the highest values in all core evaluation metrics - accuracy (0.894), recall (0.894), precision (0.886), and F1 score (F1 Score, 0.886). This result strongly proves that the NWPSO-SVM model has excellent comprehensive ability in capturing complex compliance feature patterns and achieving accurate classification prediction. Its prediction results not only have high accuracy, but also cover comprehensively, with low misjudgment rate, significantly improving the overall efficiency of compliance service pattern recognition.

In summary, the support vector machine (NWPSO-SVM) model optimized based on nonlinear weighted particle swarm algorithm proposed in this study significantly and comprehensively outperforms mainstream machine learning models such as decision trees, random forests, standard support vector machines, and BP neural networks in the key task of selecting enterprise compliance service modes. This model, optimized through intelligent algorithms, effectively overcomes the challenges faced by traditional models in processing high-dimensional and nonlinear compliant data, achieving synchronous improvement in prediction accuracy and robustness. More importantly, this study not only validates the effectiveness of a high-performance classification model, but its core value lies in providing a data-driven and highly intelligent decision support tool for enterprises in complex compliance environments, which can more accurately identify and match their required compliance service models, thereby significantly improving the efficiency of compliance management, reducing compliance risks and costs, and providing strong methodological support and practical paths for building intelligent and precise enterprise compliance risk control systems.

Acknowledgement

This paper is a phased achievement of the 2023 National College Student Innovation and Entrepreneurship Training Program Project (Project No. 202313902003), titled Exploring the Tangible Challenges and Proactive Regulatory Pathways for Consumer Privacy Protection and Utilization in the Algorithmic Era.

Additionally, this research is supported by the Department of Education of Guangdong Province Special Innovation Project for Ordinary Colleges and Universities, titled Research on Legal Protection of Algorithmic Consumers in the Digital Economy Era (Project No. 2022WTSCX154).

References

[1] Irjala, Matti, Maria Kuosa, and Marwa Hassan. "Multi-Parameter Drone Emissions Measurements for Improved Accuracy, Compliance, and Operational Efficiency in Oil & Gas." SPE Gas & Oil Technology Showcase and Conference. SPE, 2025.

DOI: 10.54254/2755-2721/2025.25153

[2] Matsumoto-Royo, Kiomi, Maria Soledad Ramírez-Montoya, and Leonardo David Glasserman-Morales. "Lifelong learning and metacognition in the assessment of pre-service teachers in practice-based teacher education." Frontiers in Education. Vol. 7. Frontiers Media SA, 2022.

[3] Elumilade, Oluwafunmike O., et al. "Optimizing corporate tax strategies and transfer pricing policies to improve financial efficiency and compliance." Journal of Advance Multidisciplinary Research 1.2 (2022): 28-38.

[4] C R K V , A J C R V V , Michel O .Risk Adjustment in Health Insurance Markets: Do Not Overlook the Real Healthy. [J].Medical care, 2023,

[5] Siegfried G , Juliane T , Stefanie S , et al.Decreasing COPD-related incidences and hospital admissions in a German health insurance population [J].Scientific Reports, 2023, 13(1): 21293-21293.

[6] Vicky V , L J W , Michelle K .Experiences of low-income college students in selection of health insurance, access, and quality of care. [J].Journal of American college health : J of ACH, 2023, 11-10.

[7] S. A M , Katia B , Michelle K , et al.Access to Specialized Care Across the Lifespan in Tetralogy of Fallot [J].CJC Pediatric and Congenital Heart Disease, 2023, 2(6PA): 267-282.

[8] Jungtaek L .Effects of private health insurance on healthcare services during the MERS Pandemic: Evidence from Korea [J].Heliyon, 2023, 9(12): e22241-e22241.

[9] Jiang, John Xuefeng, et al. "Factors associated with compliance to the hospital price transparency final rule: a national landscape study." Journal of general internal medicine (2022): 1-8.

[10] Wang, Yuqing, and Xiao Yang. "Machine Learning-Based Cloud Computing Compliance Process Automation." arXiv preprint arXiv: 2502.16344 (2025).