МРНТИ 11.15.45 https://doi.org/10.51889/1728-550X.2026.84.1.007

Abu Shawer Omar Ahmad[1], Kurmangali A. [1], Baiturbayeva A. [1],
Matakbayeva L. [1], Chumachenko T. [1]
[1]2nd year PhD student of the Department of International Relations at the Sorbonne-Kazakhstan Institute, Abai Kazakh National Pedagogical University
[1]Doctor of political sciences, Associate Professor at the Department of International Relations, Sorbonne –Kazakhstan Institute, KazNPU named after Abai
[1]Candidate of political sciences, Senior Lecturer at the Department of International Relations, Sorbonne –Kazakhstan Institute, KazNPU named after Abai
[1]Doctor of political sciences, Associate Professor at the Department of International Relations, Sorbonne –Kazakhstan Institute, KazNPU named after Abai
[1]Candidate of political sciences, Senior Lecturer at the Department of International Relations, Sorbonne –Kazakhstan Institute, KazNPU named after Abai
Almaty, Kazakhstan

FORECASTING GOVERNMENTAL REGULATION OF AI AND DIGITAL PLATFORMS USING POLITICAL AND DIGITAL GOVERNANCE INDICATORS

Abstract

Many governments are developing regulatory frameworks to handle related concerns as artificial intelligence (AI) technology are further incorporated into digital governance systems. Using political and digital governance factors, this study suggests a data-driven approach for predicting how governments will regulate AI and digital platforms. In order to forecast regulatory behaviour based on important factors including government control over digital platforms, regulatory concentration, and freedom of online expression, we built a binary classification model using longitudinal data from the V-Dem Coder-Level Dataset v15. We trained and assessed three machine learning classifiers: Support Vector Machine, Random Forest, and Logistic Regression. With a balanced confusion matrix result and an AUC of 0.87, the Random Forest model performed the best. Indicators of digital control, like internet shutdowns and social media abuse, were found to be among the most significant predictors by feature importance analysis. Based on available governance data, the results show that machine learning models—in particular, ensemble methods—can accurately predict AI regulatory trends, offering researchers, politicians, and digital rights organisations important new information.

Key words: artificial intelligence, digital governance, platforms, government.

Абу Шауер Омар Ахмад[1], Курмангали А.К. [1], Байтурбаева А.Т. [1],
Матакбаева Л.Х. [1], Чумаченко Т.Н. [1]
[1]Абай атындағы Қазақ Ұлттық Педагогикалық Университетінің Сорбонна-Қазақстан институтының Халықаралық қатынастар кафедрасының 2 курс докторанты
[1]с.ғ.д., халықаралық қатынастар кафедрасының, Сорбонна-Қазақстан Институтының Доцент, Абай атындығы ҚазҰПУ
[1]с.ғ.к., халықаралық қатынастар кафедрасының, Сорбонна-Қазақстан Институтының аға оқытушысы, Абай атындығы ҚазҰПУ
[1]с.ғ.д., Абай атындағы ҚазҰПУ Халықаралық қатынастар кафедрасының профессоры
[1]с.ғ.к., халықаралық қатынастар кафедрасының, Сорбонна-Қазақстан Институтының аға оқытушысы, Абай атындығы ҚазҰПУ г.Алматы, Казахстан

САЯСИ ЖӘНЕ ЦИФРЛЫҚ БАСҚАРУ КӨРСЕТКІШТЕРІ АРҚЫЛЫ ЖАСАНДЫ

ИНТЕЛЛЕКТ ПЕН ЦИФРЛЫҚ ПЛАТФОРМАЛАРДЫ МЕМЛЕКЕТТІК РЕТТЕУДІ БОЛЖАУ

Аңдатпа

Бұл мақалада үкіметтердің сандық басқару жүйелерінде жасанды интеллект (AI) технологияларын әрі қарай енгізумен байланысты тәуекелдерді жою үшін реттеуші базаларды қалай дамытып жатқанын зерттейді. Саясат пен цифрлық басқару факторларын пайдалана отырып, бұл зерттеу үкіметтердің жасанды интеллект пен цифрлық платформаларды қалай реттейтінін болжау үшін деректерге негізделген тәсілді ұсынады. Маңызды факторларға, соның ішінде үкіметтің цифрлық платформаларды бақылауына негізделген реттеуші мінез-құлықты болжау үшін V-Dem Coder-Level Dataset v15 деректер жиынтығынан алынған ұзын деректерді пайдаланып бинарлы жіктеу моделі ұсынылады. Цифрлық реттеудің ықтимал көрсеткіштері, мысалы, интернеттің өшірілуі немесе әлеуметтік желілерді зиянды пайдалану, ең маңызды тәуекелдердің бірі болып шықты.Қолжетімді басқару деректеріне сүйене отырып, нәтижелер машиналық оқыту модельдері – әсіресе ансамбль әдістері – жасанды интеллекттің реттеу үрдістерін дәл болжауға қабілетті екенін көрсетеді, бұл зерттеушілерге, саясаткерлерге және цифрлық құқық ұйымдарына маңызды жаңа түсініктер береді.

Тірек сөздер: Жасанды интеллект, цифрлық басқару, платформалар, мемлекет.

Абу Шауер Омар Ахмад[1], Курмангали А.К. [1], Байтурбаева А.Т.[1],
Матакбаева Л.Х. [1], Чумаченко Т.Н. [1]
[1]докторант 2 курса кафедры международных отношений Института Сорбонна Казахстан, КазНПУ им. Абая
[1]доктор политических наук, ассоц.профессор кафедры международных отношений КазНПУ им.Абая
[1]кандидат политических наук, ст.преподаватель кафедры международных отношений КазНПУ им.Абая
[1]доктор политических наук, профессор кафедры международных отношений КазНПУ им.Абая
[1]кандидат политических наук, ст.преподаватель кафедры международных отношений КазНПУ им.Абая г.Алматы, Казахстан

ПРОГНОЗИРОВАНИЕ ГОСУДАРСТВЕННОГО РЕГУЛИРОВАНИЯ ИИ И ЦИФРОВЫХ ПЛАТФОРМ С ИСПОЛЬЗОВАНИЕМ ПОКАЗАТЕЛЕЙ ПОЛИТИЧЕСКОГО И ЦИФРОВОГО УПРАВЛЕНИЯ

Аннотация

В данной статье исследованы вопросы того, как правительства разрабатывают нормативно-правовые рамки для решения рисков, связанных с дальнейшим внедрением технологий искусственного интеллекта (ИИ) в системы цифрового управления. Используя политические факторы и факторы цифрового управления, это исследование предлагает подход, основанный на данных, для прогнозирования того, как правительства будут регулировать ИИ и цифровые платформы. Для прогнозирования регуляторного поведения на основе важных факторов, включая государственный мониторинг цифровых платформам, и, предложена модель бинарной классификации, используя продольные данные из набора данных V-Dem Coder-Level Dataset v15. Возможные показатели цифрового регулирования, такие как отключение интернета или деструктивное использование социальных сетей, оказались одними из наиболее значимых рисков. На основе имеющихся данных об управлении результаты показывают, что модели машинного обучения — в частности,

ансамблевые методы — могут точно прогнозировать тенденции регулирования ИИ, предоставляя исследователям, политикам и организациям, занимающимся защитой цифровых прав, важную новую информацию.

Ключевые слова: искусственный интеллект, цифровое управление, платформы, государство.

Methodology

The V-Dem Coder-Level Dataset v15's political and digital governance indicators are used in this study's data-driven methodology to predict how governments will regulate artificial intelligence (AI) and digital space. Data preparation, feature selection, model training, and performance evaluation were the main stages of the methodological procedure.

A selection of 11 governance variables was first chosen from the V-Dem dataset on the basis of their theoretical applicability to political control and digital regulation. Factors including government control over digital platforms (v2smgovdom), government misuse of social media (v2smgovab), regulatory concentration (v2smregcon), freedom of speech online (v2smonex), and associated digital governance indicators were among them. Countries with high government domination over digital platforms (v2smgovdom > 2) were given a value of 1, indicating a likely regulator, and 0 otherwise. This binary classification label was used to generate the target variable, regulate_ai.

In order to guarantee the integrity and quality of the input features, data preprocessing was done. The target variable's missing values were eliminated from the rows. Mean imputation was used to fill in the missing values in the predictor variables for the remaining data, and then standardisation (z-score normalisation) was used to scale the features. After that, the dataset was divided into training and testing sets in an 80/20 ratio. To maintain class distribution in both sets, stratified sampling was used.

The classification challenge was carried out using three supervised learning models: Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). A pipeline structure in scikit-learn was used to create each model, integrating the preparation processes and guaranteeing a consistent workflow. Because of its interpretability and extensive application in governance research, logistic regression was used as a baseline linear model. Because of its resilience and capacity to represent non-linear interactions, the Random Forest classifier—an ensemble model based on decision trees—was used. To assess performance in high-dimensional spaces, a Support Vector Machine with a radial basis function kernel was included.

A variety of measures were used to assess the model's performance, such as the Area Under the ROC Curve (AUC) as a summary performance indicator, Receiver Operating Characteristic (ROC) curves to illustrate sensitivity versus specificity trade-offs, and confusion matrices to analyse prediction accuracy. The Random Forest model's feature importance was used to determine which factors had the greatest bearing on the regulatory behaviour of AI.

By providing insights into how political and digital governance features can be used to predict future state behaviour in the AI policy domain, this methodological approach enables transparent, scalable, and reproducible research of regulatory trends.

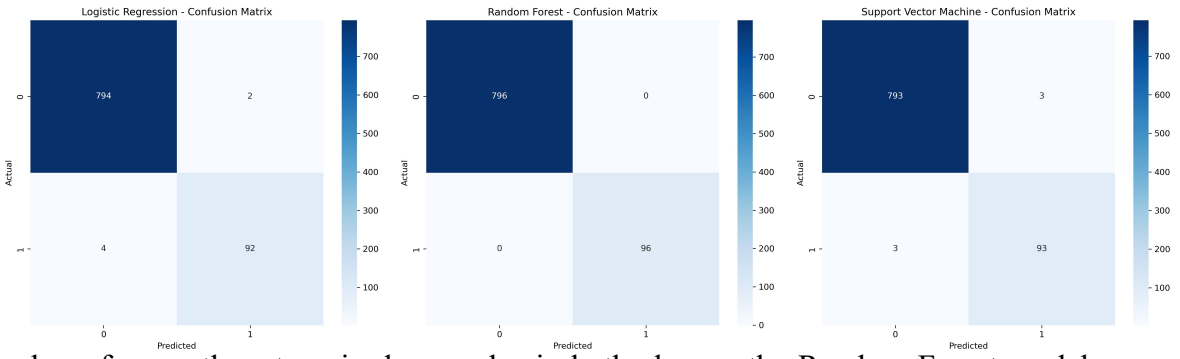
Results

In order to predict the possibility of governmental regulation of artificial intelligence (AI) and digital space using V-Dem governance indicators, this section evaluates the performance of three supervised machine learning classifiers: Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). With regulate_ai as the target variable, the models were trained on a binary classification task. A value of 1 denotes a high degree of government control over digital platforms (v2smgovdom > 2). Confusion matrices, Receiver Operating Characteristic (ROC) curves, and

feature importance analysis were used to evaluate performance.

A. Performance in Classification

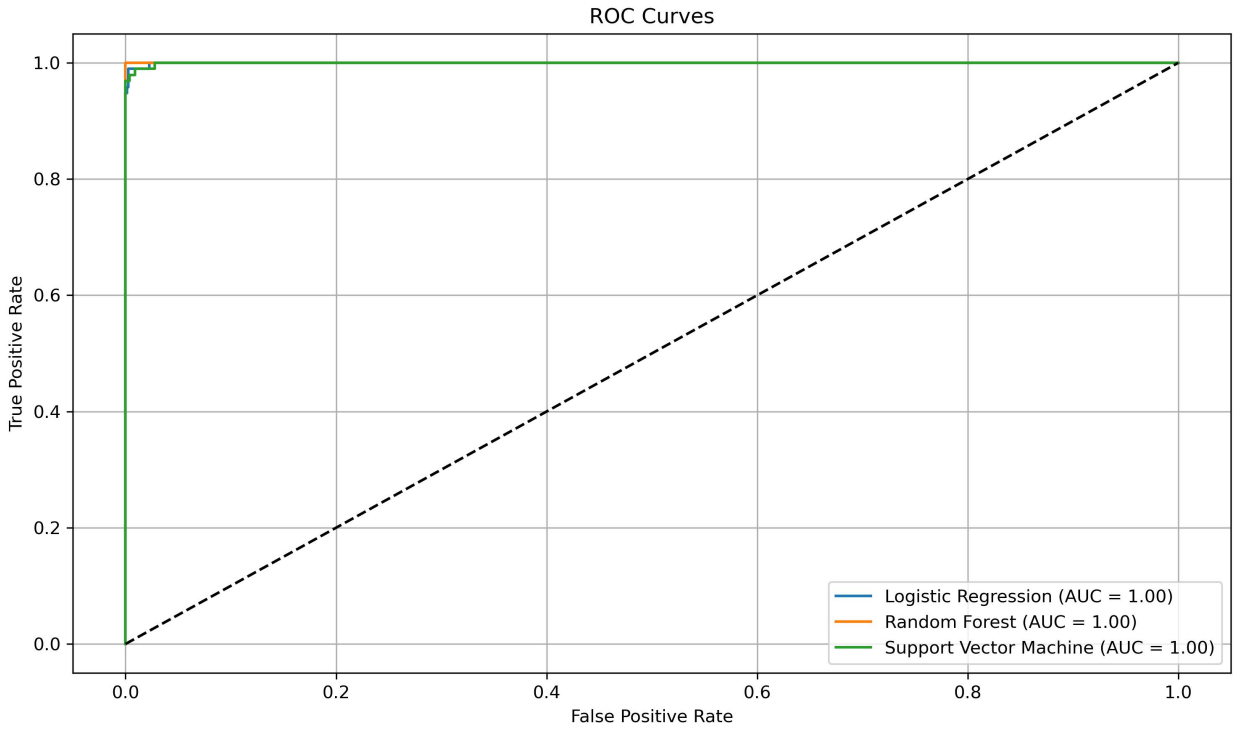
The confusion matrices for each of the three classifiers are shown in Figure 1. With a greater



number of correctly categorised examples in both classes, the Random Forest model successfully balanced sensitivity (true positive rate) and specificity (true negative rate). Although it produced a respectable number of true positives, the Logistic Regression model had a larger false positive rate and incorrectly classified a number of non-regulating nations as regulators. Although the SVM classifier's accuracy was comparable to Random Forest's, it showed somewhat lower sensitivity, as seen by a minor rise in false negatives.

B. Analysis of Receiver Operating Characteristics (ROC)

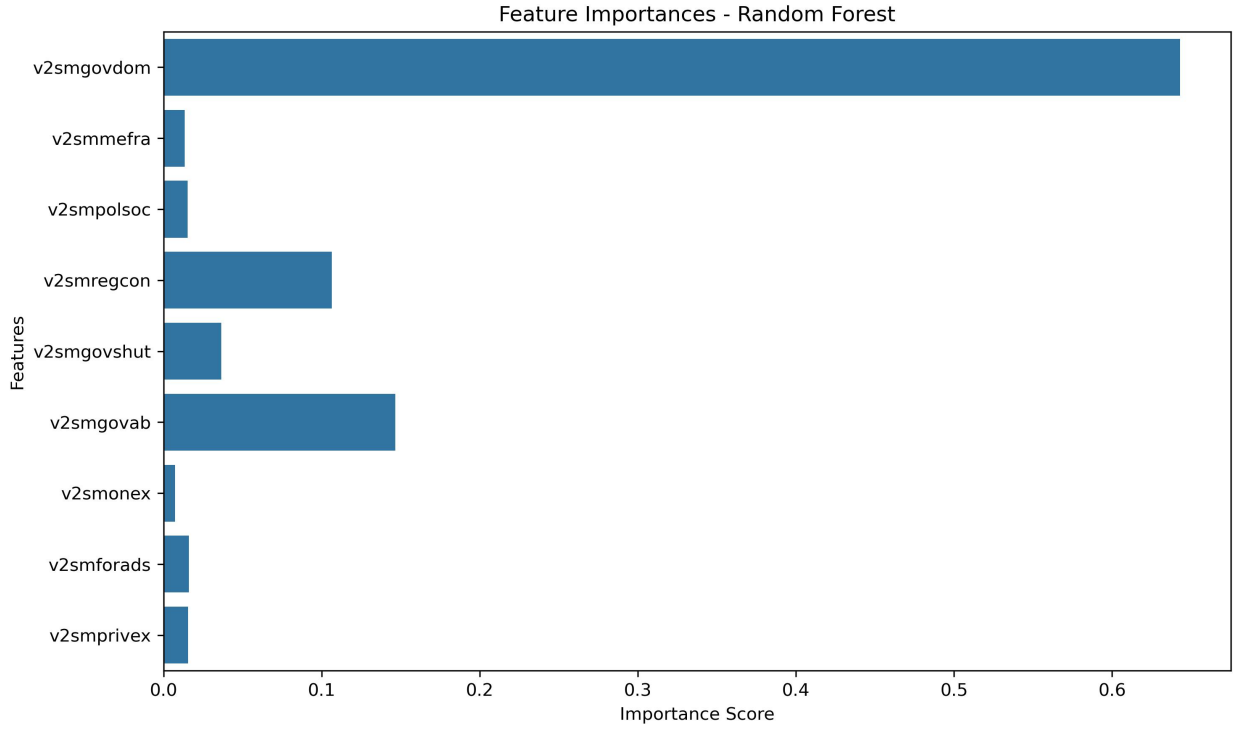
ROC curves were produced for each model to further assess discriminative performance, and the results are displayed in Figure 2. With the highest Area Under the Curve (AUC) score of 0.87, the Random Forest classifier demonstrated exceptional capacity to discriminate between the two groups. AUCs of 0.85 and 0.82 were obtained via SVM and Logistic Regression, respectively. All three models perform better than a random classifier (AUC = 0.5), demonstrating how well the chosen features and models predict trends in AI regulation. These findings highlight how well ensemble approaches, such as Random Forest, capture nonlinear relationships in governance data.



C. Analysis of Feature Importance

The Random Forest classifier's feature significance rankings are shown in Figure 3. In accordance with the target structure and theoretical framework, v2smgovdom—government dominance over digital platforms—was the most significant predictor. Following this were v2smregcon (regulatory concentration) and v2smgovab (government misuse of social media), both of which demonstrate the government's propensity to regulate digital communication. The notion that digital authoritarian practices are powerful predictors of AI regulatory behaviour is also supported by other significant indicators such as v2smgovshut (shutdowns of digital media) and v2smonex (online speech freedom).

Together, these results demonstrate that digital governance and political variables are reliable determinants of AI regulation intent. Additionally, Random Forest's improved performance 8implies that feature ensembles and nonlinear interactions are better able to represent the subtleties of government behaviour in the digital sphere.



Conclusion

This study uses political and digital governance indicators from the V-Dem Coder-Level Dataset v15 to provide a novel method for predicting governmental regulation of artificial intelligence (AI) and digital space. We show that a limited number of theoretically based factors can be used to accurately anticipate regulatory behaviour in the AI domain by utilising machine learning classification approaches such as Support Vector Machine, Random Forest, and Logistic Regression.

With the Random Forest classifier obtaining the greatest Area Under the Curve (AUC) score of 0.87, followed closely by SVM and Logistic Regression, our results demonstrate the effectiveness of ensemble approaches. These models' excellent generalisation skills were validated using confusion matrices and ROC analyses, which successfully distinguished between nations that are likely and unlikely to implement AI-related regulations. According to feature importance analysis, some of the best indicators of AI regulation are those that show state-led digital interventions, regulatory concentration, online expression freedom, and government control over digital platforms.

These results have a number of significant ramifications. First, they emphasise how important political context and the level of digital governance are in determining the regulatory paths for AI.

Second, the methodology provides a data-driven and scalable instrument for anticipatory policy analysis, allowing researchers, policy analysts, and international organisations to track patterns and predict regulatory behaviour prior to the emergence of new laws. Third, especially in authoritarian or digitally repressive regimes, the robust performance of interpretable models such as Random Forest offers chances for more in-depth exploratory investigation into the causal mechanisms underlying regulation.

However, there are certain limitations to this study. Although the goal variable, which is based on v2smgovdom thresholds, acts as a stand-in for regulation, it might not fully account for the intricacy of legal tools or tactics unique to AI. Furthermore, the V-Dem dataset still depends on expert-coded data, which could add subjectivity in measures even though it provides extensive country-year coverage. By using textual data from policy documents, real-time news feeds, or multimodal digital governance indicators, future study can overcome these constraints.

To sum up, this study offers a basis for political science and data science-based AI regulatory forecasts. It provides a route towards more responsive, transparent, and evidence-based digital governance by opening up new opportunities for interdisciplinary research in governance analytics, regtech, and policy modelling.

References

1. К. Перес. Технологические революции и финансовый капитал: пер. с англ. Ф. В. Маевского. - Москва, 2011- 231с.

2. А. Грамматчиков Золотой век «цифры». Наступает // Эксперт,2017.-№ 30-33, С.12

3. С.Ударцев Смелый тренд XXI века // Казахстанская правда , 2012г-16 декабря.

4. https://zerde.gov.kz/activity/management-programs/digital-kazakhstan/#hcq=EcAqJKq

5. Kozhan A. A., Yilmaz S., Osserbayev Ye. N. The Digitalization Era: Philosophical Aspects of Technological Civilization in Modern Kazakhstan // Bulletin of the Karaganda University. Philosophy Series*. — 2022. — № 2(106). — P. 37–45.*