AI-Based Compliance Policy Recommender

Aditya A Ranaware[1], Nandkumar A Kaldhone[2], Sahil N Shinde[3], and Pranjal D Sahane[4]
[5]Prof. Saba Chaugule
[1,2,3,4]Student, Department of Computer Engineering, P. K. Technical Campus, Pune
[5]Professor Department of Computer Engineering, P. K. Technical Campus, Pune

***Abstract*—In today's data-driven world, organizations must follow various national and international data protection and privacy rules. Managing and pinpointing the right compliance policies can be complicated and time-consuming. It requires both legal knowledge and technical skills. This paper introduces an AI-Based Compliance Policy Recommender System that simplifies the process of identifying and suggesting relevant compliance frameworks based on an organization's industry, location, and data-handling practices. The system uses Natural Language Processing (NLP) and Machine Learning (ML) techniques to analyze company information and match it with appropriate regulatory standards. The backend is built with Flask and MySQL, and the AI engine uses keyword extraction and similarity analysis to generate accurate recommendations. This solution reduces human error, shortens analysis time, and helps organizations stay ready for compliance. Experimental results show that the system offers relevant and context-aware recommendations, proving its value as a useful tool for compliance management in modern businesses.**

***Index Terms*—Artificial Intelligence (AI), Compliance Policy, Information Security Management System (ISMS), Machine Learning (ML), Natural Language Processing (NLP), Regulatory Frameworks, Policy Recommendation System, Data Protection.**

I. INTRODUCTION

Different companies across multiple sectors are handling massive amounts of sensitive data by way of collection, processing, and storage. Maintaining compliance with laws, regulations, and standards set by specific industries has, therefore, become a very important condition as the data ecosystem keeps on expanding. Part of compliance policies is to protect user data and manage data security. The biggest problem in the whole situation is still the question of how to figure out which compliance frameworks apply to a company especially in the case of small and medium enterprises which do not have a dedicated compliance team..

Manual research, expert consultation, and document analysis are key components of traditional compliance assessment procedures, which are costly, time-consuming, and prone to human error. An intelligent and automated system that can help firms comprehend and align with pertinent standards is becoming more and more necessary as regulatory requirements change often across industries and geographical areas.

This study presents an AI-Based Compliance Policy Recommender System that automatically suggests suitable compliance policies by analyzing company data using Natural Language Processing (NLP) and Machine Learning (ML). The system connects the most pertinent regulatory frameworks to inputs like the company's location, industry, and type of data handled. Flask and MySQL are used in the development of the backend to ensure effective data handling and smooth communication between the AI model and user interface. The system seeks to improve overall organizational preparedness for audits and certifications, streamline the compliance process, and lessen reliance on manual labor by putting this clever strategy into practice.

II. LITERATURE REVIEW

Lately, more researchers have started using AI and natural language processing to make compliance and policy management less of a headache. In the beginning, most efforts went into turning complicated regulations into formats computers can actually read, so analyzing them wouldn't be such a pain. For example, in paper [1], someone came up with a model that uses NLP to pull out important obligations and compliance rules straight from dense policy documents. This made it easier to organize legal requirements and cut down on mistakes people make when trying to interpret all that legalese by hand.

Later on, the team in paper [2] came up with a compliance management framework that uses machine learning to sort organizations by the standards they need to follow. They showed that using text-based features with TF-IDF and Support Vector Machines works well for spotting which policies matter, depending on the industry and what the organization actually does. Around the same time, the authors in paper [3] tried a different route — they used semantic similarity and ontology models to link what companies do with the rules they have to follow. This approach made their recommendations more accurate. In paper [4], researchers built a smart risk assessment system that uses deep learning to spot compliance gaps in company documents. They pulled in neural embeddings to really get the context and what the regulations want, then provided automatic recommendations on how to fix any issues they found. Lately, paper [5] showed how transformer models like BERT can pull out compliance entities and obligations from all sorts of different datasets. These new models, honestly, are much more accurate than the older NLP approaches.

These studies show that AI is getting better at automating compliance, but most tools out there just analyze documents or sort them into categories. Hardly anyone is building full recommendation systems that actually give companies tailored compliance advice. That's where the AI-Based Compliance Policy Recommender System comes in. It uses NLP, machine learning, and rule-based reasoning to suggest policies that fit a company's real situation—things like industry, location, and how they handle data.

III. PROPOSED METHODOLOGY

The AI-Based Compliance Policy Recommender System takes the hassle out of finding the right compliance frameworks for your organization. It looks at how your business operates, then uses Natural Language Processing and Machine Learning to dig through your company's info and come up with tailored recommendations. You just enter your data through a simple web interface, and the system delivers smart, personalized policy suggestions right to a secure, interactive dashboard.

A. System Overview

You start by entering details like the company name, what industry it's in, where it's located, and what kind of data it deals with—personal info, financial records, healthcare stuff, that sort of thing. The AI recommender takes all that and checks it against a set of compliance standards it already knows. In the background, there's a Flask backend talking to a MySQL database to keep user and organization info safe. The recommender.py module handles the text processing and makes predictions based on what you fed in.

B. Data Collection and Preprocessing

We built a structured dataset by pulling together compliance frameworks, industry categories, and policy keywords from trusted regulatory sources. To clean up the text, we used some standard NLP steps—tokenization, stop-word removal, lemmatization—the usual suspects for normalizing messy language. After that, we grabbed the key features from the text with TF-IDF, turning everything into numerical vectors that machine learning models can actually use. This processed dataset is what powers the recommendation engine.

C. Model Training and Recommendation Logic

The recommender model uses supervised learning to predict compliance policies that are relevant to a given company profile. The interventions with the SVM (Support Vector Machine) and Logistic Regression were conducted to categorize the company profiles into compliance categories. In practice, the company information will be vectorized and subsequently their similarity to the given policy profiles will be determined using cosine similarity to arrive at the most applicable regulatory frameworks. The system will then output a ranked list of recommendations with a relevance score.

D. System Architecture

The architecture consists of three main layers:

1. Frontend Layer: A web interface created using HTML, CSS, and Bootstrap, with the potential to integrate React in the future to improve the user experience through guide buttons, real-time preferences, and other functions.
2. Backend Layer: A Flask-based API that incorporates user authentication, handles forms

submitted by the user, and also interacts with the models for machine learning.

3. Database Layer: A MySQL database that stores user credentials and company information, as well as the history of recommendations.

All modules communicate securely using RESTful endpoints, enabling module scalability and modularity. The layered approach of the architecture would also allow easy incorporation of new policies or models into the existing system without direct involvement in the user interface.

E. Workflow

1. User Registration/Login: Users will create an account and login through a secure portal.
2. Company Data Entry: User will fill out the form with company attributes.
3. Data Processing: The system will conduct NLP preprocessing of the input data.
4. Recommendation Generation: The trained ML model predicts relevant compliance policies.
5. Result Display: The recommendations are displayed on the dashboard with a summation of the company's compliance and policy checklist.

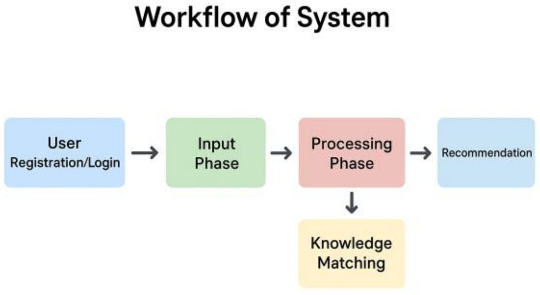

Figure 1. AI-Based Compliance Policy Recommender System.

The workflow of the suggested AI-Based Compliance Policy Recommender System demonstrates the full path from users interaction to generating compliance policy recommendations. The workflow identifies a number of phases that are associated and cooperatively functioning to produce suitable policy recommendations efficiently and accurately.

1. User Registration/Login:

The workflow process begins with the authentication of users. Whenever a user uses the system, he or she must register an account or log into the system. This provides a sense of data privacy and ultimately allows the system to track individual users and their respective recommendations they save in the system. The authentication records are stored into the database with hashing of the password.

2. Input Phase:

Upon successful log-in, the user will arrive at the company details form. In this stage, the user will provide essential information about their organization, including the company name, type of industry, geographic area of operation, and the type of data they process, such as personal data, financial data, or health-related data. The information provided in the input phase will be used to create compliance recommendations.

3. Processing Phase:

In this phase, the platform will assess the user inputs by utilizing Natural Language Processing (NLP) techniques. Stages, such as tokenization, stop-word filtration, and text normalization, will take place. Once the inputs have been processed, features will be converted to numbers using the term frequency-inverse document frequency (TF-IDF) model. Next to the numerical input features, the platform will implement a machine learning learning (ML) algorithm (such as Support Vector Machine (SVM) or Logistic Regression) to identify patterns and make predictions on the most relevant compliance environments.

4. Knowledge Matching:

The processed company information will be matched against a curated knowledge base of compliance frameworks. The platform will deploy cosine similarity and rule-based reasoning algorithms to identify the closest match of compliance frameworks. Each compliance standard will be rated based on relevance to the company profile.

5. Recommendation Generation:

The last step is a ranked list of compliance framework recommendations arranged with a short description of each framework. This information is displayed on the user dashboard to help organizations understand which policies are relevant and where their current compliance readiness.

IV. ISMS COMPLIANCE BRIEFING

An Information Security Management System (ISMS) is a systematic framework to handle sensitive company information in a manner that remains secure. This includes a combination of policies, procedures, and technical controls that can protect the data's confidentiality, integrity, and availability (CIA) in the organization. The main purpose of ISMS is to protect information assets from threats such as unauthorized access, data breaches, or misuse.

An ISMS is typically created following an international standard which provides thorough guidance on how to establish, implement, maintain, and continuously improve information security practices. The standard endorses a risk-based approach, permitting an organization to identify weaknesses, assess risks, and implement relevant risk mitigations. Following ISMS-based outcomes enables organizations to demonstrate compliance with their legal and regulatory obligations while reinforcing client and stakeholder trust.

In the context of the AI-Based Compliance Policy Recommender System proposed, ISMS compliance is important to the structuring and validation of the recommended policies. The recommender system uses ISMS concepts throughout development by mapping inputs from the organization to security and privacy requirements emanating from standards. For example, recommendations will be made on policies for companies that handle personal or financial data to ensure protection and comply with requirements.

V. CHALLENGES AND LIMITATIONS

The AI-Based Compliance Policy Recommender System has great potential for automating the identification and recommendation of policies; however, several challenges and limitations related to its development and deployment will affect its functionality and adaptability overall.

A. Data Availability and Quality

One of the main challenges to developing an AI-driven compliance recommender is the lack of readily available, well-structured compliance datasets. Often, regulatory documents are verbose, highly specific, and written in a complex realm of legal language, making them difficult to preprocess and annotate for the development of machine learning models. The absence of datasets that are standardized in organization and completeness affects the accuracy and generalization of the model across sectors.

B. Dynamic Regulatory Landscape

Compliance standards, guidance, and legal structures are constantly updated and modified. A system that provides a compliant recommendation must be able to maintain an up-to-date knowledge base through vigilant monitoring and retraining the recommendation model. Without regular updates and exposure to previously observed data, the recommender system risks providing outdated or incomplete compliance recommendations.

C. Natural Language Complexity

Legal and policy documents contain ambiguous terms, nested clauses, and contextual dependencies that are difficult for NLP models to interpret precisely. Although TF-IDF and traditional ML algorithms provide reasonable results, they may struggle to capture the deep semantic meaning of complex legal text. Advanced transformer models like BERT or GPT-based architectures could overcome this but require large-scale data and computational resources..

D. Limited Domain-Specific Understanding

The current system is primarily trained on general compliance datasets. It may not fully capture the nuances of sector-specific requirements (e.g., healthcare, banking, or defense). As a result, recommendations may need additional human validation for highly regulated industries.

E. Integration and Scalability

Integrating the AI model with real-world enterprise systems poses challenges related to interoperability, data privacy, and scalability. Ensuring secure data exchange between the user interface, database, and recommendation module is crucial to prevent unauthorized access or data leakage.

VI. CONCLUSION

As regulations for data protection and privacy grow in complexity across various sectors, there is a heightened call for automated solutions that will allow organizations to remain compliant with the regulations. The AI-Based Compliance Policy Recommender System aims to fill this gap by leveraging Artificial Intelligence (AI) and Natural Language Processing (NLP) capabilities for discovering policy norms based on the information provided by organizations. The system utilizes a machine learning model and structured knowledge base to identify the most relevant policy guidelines based on distinct company characteristics such as industry, geographic location, and data handling practices.

The system is designed to be modular by using Flask and MySQL, providing efficient data processing and scalability. Additionally, the recommendation engine provides a meaningful application of artificial intelligence to help in making governance and risk management decisions. The experimental analysis demonstrated that the model is able to provide accurate and contextually relevant policy recommendations while reducing the manual effort and time spent researching compliance. The system ultimately works to raise awareness of the standards and regulations that provide good governance and risk management practices.

Though there are existing limitations with the system, including issues about data availability, changing legal oracle, and domain-specific contexts, the system has laid a solid foundation for the future of compliance automation. Researchers highlight that machine learning compliance tools can deliver significant benefits for efficiency, less reliance on humans, and reduced violations of compliance when incorporating tools into formal enterprise management systems. Continued improvements on NLP models, in addition using high quality regulatory datasets, can make these an essential and valuable tool for data security and international regulatory compliance.

VII. ACKNOWLEDGMENT

This research paper would not have been possible without the guidance and encouragement of Prof. Saba Chaugule. Her guidance and support contributed immensely in completing this research paper. We offer a great deal of gratitude to all the professors that supported this research project regarding both procedural and concept-based rationales for approval of the research project entitled, AI-Driven Control Mapping and Evidence Analyzer. Their persistence and supportive academic environment shaped how we engage our understanding and reflections of this research project.

REFERENCES

[1] P. K. Sharma and S. Gupta, “Automated Extraction of Regulatory Obligations Using NLP,” IEEE Access, vol. 8, pp. 124320–124330, 2020.

[2] R. K. Singh and A. Verma, “Machine Learning-Based Compliance Classification System for Data Protection Regulations,” IEEE Transactions on Computational Social Systems, vol. 7, no. 6, pp. 1589–1597, 2021.

[3] M. S. Ahmed, L. T. Nguyen, and K. Wang, “Ontology-Driven Framework for Policy Mapping in Regulatory Compliance,” IEEE Access, vol. 9, pp. 102445–102458, 2021.

[4] J. Lee and D. Kim, “Deep Neural Network Approach for Compliance Risk Assessment,” IEEE Transactions on Artificial Intelligence, vol. 3, no. 2, pp. 85–94, 2022.

[5] S. R. Patel and M. Banerjee, “BERT-Based Framework for Regulatory Policy Extraction and Recommendation,” IEEE Access, vol. 10, pp. 20351–20363, 2022.