Exploring HCP Influence Through Machine Learning: A Predictive Analysis of Lyumjev Prescribing Trends in the North Zone in India

Gourav
Assistant Professor in Computer Science
PCTE Institute of Engineering & Technology, Punjab, India

Shairy
Assistant Professor in Computer Science
PCTE Institute of Engineering & Technology, Punjab, India

Kamalpreet Kaur
Assistant Professor in Computer Science
PCTE Institute of Engineering & Technology, Punjab, India

ABSTRACT
The understanding of the prescribing conduct of the Healthcare Professionals (HCPs) and segmentation over the adoption and the use of medicinal drugs. This study explores the effects of Healthcare professionals on the prescription patterns of the Lyumjev a quick-acting insulin variant, in Northen Region in India. The study employs machine learning methods and models to forecast Lyumjev prescription behavior, examining the fundamental components of the influence of HCP decision-making [6]. By analyzing prescription data, patient demographics, HCP specialties, and clinical recommendations, the study aims to find the connections between each of these factors and the likelihood of prescribing Lyumjev.

The research and the study leverage the machine learning models, and algorithms, including decision tree, random forest classifier (RFC), and support vector machine (SVM) to make the predictive model. These models are trained and tested by using real-life data which is filled by the medical representatives while making the deliveries of the Lyumjev to the North region doctors. The prescribing data can make the future trends of the HCPs influence. Distribution is identified as a significant predictor of prescribing behavior [3]. The study also explores the role of clinical guidelines and treatment protocols in shaping prescription choices.

The findings provide valuable insights into how HCP characteristics influence their prescribing habits and the factors driving the adoption of Lyumjev over other insulin analogs. Furthermore, the paper provides the predictive abilities of the machine learning models and the methodologies not only in the healthcare sector but also in the marketing of medicines. The findings have practical relevance and the results for the pharmaceutical companies where they can refine their marketing as well as HCPs to influence.

Keywords
HCPs, Healthcare Professionals, Insulin, Lyumjev, Machine Learning, Prediction, Predictive behavior, Prescribing Behavior, North Zone, SVM, RFC, decision making, decision tree, Support vector machine, random forest classification, classifiers

1. INTRODUCTION
The prescribing behavior of healthcare professionals (HCPs) plays an important role in the adoption and use of medicines or pharmaceutical products. In the last year, numerous researches and advancements have been done in diabetic medicines one of the medicines is Lyumjev rapid action insulin by one of the pharmaceutical company has been used and marketed by the medical representatives to the doctors for the Type-1 and Type-2 diabetic patients.

This paper aims and provide the underlying factors that influence HCPs' decision-making process in prescribing the Lyumjev medicine to the doctors to the patients in the north zone of India. The north zone of India, which encompasses a diverse population of patients and healthcare, is for studying the prescribing trends of the medicine and the patterns of the HCPs influenced by the complex set of labels including the medicines of the other companies and the adoption rate of the other medicine of the same genre.

To analyze the large dataset and patterns of the HCPs influence this research takes the help of machine learning models and the methodologies for training and testing the models to make the predictions over the influence. By using these models this study seeks to predict prescribing behaviors and identify the various factors that influence the adoption of this medicine [6].

This paper provides the results of the medicine in the northern zone of India. The following sections of this paper contain the methodology, objectives, data patterns, results of the model, and the decisions of the implications and the future of the paper over the other zones of India.

2. LITERATURE REVIEW
The interaction between the HCPs and the prescribers of medicine has generated significant attention over the years. This impacts the influence of the HCPs over the prescribing behavior of the doctor concerning their drugs [5].

Yimenu et al. explored the attitudes, influence, and acceptance of drug promotion by the HCPs among doctors in Ethiopia [4]. The was conducted in Bahir Dor and Gondar cities, which highlights the influence of pharmaceutical promotions. It emphasized that the doctors and the HCPs are shaped by population size, healthcare access, and the professional dynamic [4]. This study provides the results of drug promotion, in the particular region where the resources are limited.

Khazzaka investigates the influence of HCP's marketing strategy on physicians prescribing in Lebanon [2]. The study results highlight the significant impact of the HCP's marketing strategy over the free drug samples motivates the doctors' prescribing patterns and raises the influence rate related to the drug [2]. The study further defines the patterns of the influence of HCPs over physician's demographic and professional characteristics, such as age, gender, and the location of the practice.

Similarly, Anderson et al.examined the reliability of physicians and HCPs for prescribing decisions in the United States [3]. The study found the influence of the HCPs with 29% frequently relying on them when deciding on a new prescription [3]. Such as practice type and availability of drug samples strongly influenced this reliance. This results as the sample distribution isan important part of shaping the influence of the drug.

The studies have been conducted over the questionnaire to find the patterns of the influence of the HCPs over the doctor. To find out the prescription behaviors of the physicians.

3. METHODOLOGY

This study investigates the predictive factors influencing Lyumjev prescription patterns by physicians in the North region. Utilizing machine learning models, the research aims to analyze prescribing data to determine the extent to which Healthcare Professionals (HCPs) influence physician prescribing decisions for Lyumjev compared to other medications. Specifically, the study will examine whether a medication delivered by HCPs is subsequently prescribed by physicians to their patients and, if so, predict the influence rate of HCPs on physician prescribing behavior within the North zone.

The analysis extends beyond regional trends to examine city-level prescribing patterns, considering the influence of HCPs on physician prescribing behavior. The study investigates the impact of region, city, and area on HCP influence. Furthermore, it explores the relationship between physician specialty and Lyumjev prescription rates.

3.1 METHOD

This research leverages primary data collected by Healthcare Professionals (HCPs) employed by the company [1]. Unlike studies relying on questionnaires, this research directly analyzes HCP-collected data to predict Lyumjev prescription patterns.

A data collection sheet was designed with specific labels to guide HCPs in gathering data from various cities within the North region. The resulting dataset comprises 2216 entries, reflecting HCP influence patterns categorized into three target values: Low, Medium, and High. These values represent the predicted influence rate of HCPs on physician prescribing decisions for Lyumjev.

Machine learning models, including Decision Trees, Random Forest Classification (RFC), and Support Vector Machines (SVM), were employed to generate predictions. Model performance was evaluated using metrics such as Precision, Recall, F1-score, and Support, as derived from the Confusion Matrix, to identify the most accurate predictive model.

The Method of the models have been used to find out the results by applying the mathematical equations of the Decision Tree, Random forest classification (RFC) & Support Vector Machine (SVM) as follows:

Decision Tree: E(D)=−i=1∑classpilog2(pi)

In this, the *D* is the dataset or the group of data points that are used in this method to analyze the results of the data. *Class* is the classes that are present in the dataset or the possible outcomes or labels of the datasetby Sarker [6]. This dataset has 3 labels of the data *Low, High, and Medium*. The labels of the dataset have been used as 0,1,2 which predicts the results *as HCPs whose 1 Rxn can influence other HCPs Rxingbehavior*, where *Pi* is the proportion of the data and *log2(pi)* is the entropy function.

3.2 RESULTS

A data set has the total number of rows as 2216 and after adding the data into the sheet by the HCP the total value of the data is 1640 rows which is approximately 74% of the data which is collected by the HCP to know the behaviour of the data. On that, the multiple data pre-processing methods have been applied to make the data efficient to feed into the machine to make the predictions over the data to apply the machine models.

After applying the pre-processing method [1], the data is reduced to 70% for making the predictions over the influence as the target values where classes of the data are*low, high, and medium*. Where each class represents the *HCPs whose 1 Rxn can influence other HCPs'Rxingbehavior*, after applying the multiple methods over the data the results have been found in***Table 1***:

Table 1. Results of the Various Prediction Models

Metric	**Prediction Models**		
	Decision Tree	**Random Forest**	**Support Vector Machine (SVM)**
Accuracy	*0.9009*	*0.8784*	*0.8784*
Precision (Macro Avg)	*0.84*	*0.93*	*0.76*
Recall (Macro Avg)	*0.69*	*0.55*	*0.68*
F1-Score (Macro Avg)	*0.75*	*0.60*	*0.71*
Precision (Weighted Avg)	*0.90*	*0.89*	*0.87*
Recall (Weighted Avg)	*0.90*	*0.88*	*0.88*
F1-Score (Weighted Avg)	*0.89*	*0.86*	*0.87*

Table 1 shows the comparison of all the models that have been used to analyze the data and make predictions on the data. Decision trees have more accuracy than the other models. Where the results are calculated over the models and the confusion matrix to find the weighted average and Macro average of the data based on *Precision, Recall, and F1-score* as well.

The Decision Tree classifier produced the highest accuracy result of 90.09% and exceeded the accuracy of both Random Forest and SVM. The Decision Tree model also provided the highest scores for every weighted average metric as well: precision (0.90), recall (0.90), and F1-score (0.89). These scores indicate that the Decision Tree was successful overall, and alsoin regards to class imbalances as reflected in its weighted scores.

Although Random Forest produced the highest macro-average precision (0.93) and significantly lower macro recall (0.55) and F1 (0.60), there is a significant contrast in model accuracy. Random Forest was highly precise in predicting the individual classes but failed to capture the significant number of true positives needed for increased recall. This indicates that Random Forest had a better chance of reducing false positives but resulted in increasing false negatives.

Despite not producing a higher score than the Decision Tree, the SVM appeared balanced with high scores in macro recall (0.68) and macro F1 (0.71). The SVM micro values do indicate some variability in representing multiple classes consistently, which is reasonable because often both precision and recall are important.

The Decision Tree model stands out as the best classifier in the study in terms of high overall accuracy and individual class performance. An additional evaluation of the macro and weighted average metrics also reiterates that the Decision Tree is the best model for the dataset under assessment.

3.2.1 Analysis

According to Arora et al.,visualization focuses on the drug prescription and the doctor's specialty.[7] ***Figure 1*** depicts the kind of doctors who have prescribed Lyumjev to their patients, as provided by HCPs to doctors in India's north area. The statistics suggest that endocrinologists had a higher perception of Lyumjev than other doctors [7]. As a result, HCPs may focus on the same field doctors throughout the northern area, increasing the medicine's effect.

Figure2 highlights the distribution and variability of *Lyumjev* across based on the north zone.It provides the data visualization of HCP’s influence rate that they have done during their fieldwork. The influence rate of all the locations of the Lyumjev can be clearly from the data which was collected by the HCPs. Some of the locations have zero influencing rate and others are high. So they need to focus on the locations where the influencing rate.

On the other hand, this predicts that the areas thathave zero influencing rate either haven’t approached the doctors or the doctors did not like to prescribe the medicine to their patients or they are using any other company medicine to prescribe to their patients.

Figure 3 displays the grouped bar plot distribution over the influence of Lyumjev and the area business managers'headquarters (ABM HQ) distributions. It shows the distribution of Lyumjev values across several categories, impacted by both ABM HQ and Lyumjev. The influence of HCPs to prescribe the medicine to the patients based on their cities.The figure shows significant patterns: certain Lyumjev levels, such as 6.0, dominate certain groupings, but others, such as 10.0 or 17.0, are unusual.

Figure 4 and ***Figure 5*** depict the visualization of the influence of Lyumjev based on the Regional business managers and the Zone business managers [7]. It provides the results of the influence of the HCPs based on the regional base and the zone based on the prescription rate of the doctors related to the drug.

After analysis of the results by using the machine learning models and the Visualization of the data. From the patterns, this paper provides that knowledge about the influence of HCPs has changed the prescription behavior of the physician with their sample drugs. It emphasizes the patterns of the drug Lyumjev which is prescribed by the physicians to their patients in the north zone. The prediction rate of the data is 90% after applying the decision tree and the other are on approximately 88% prediction rate. The graphical representation tells the rate of the Influence based on the Area, Region, and Zone and how the influence of the HCPs changed the behavior of the prescription of the Lyumjev drug to their patients.

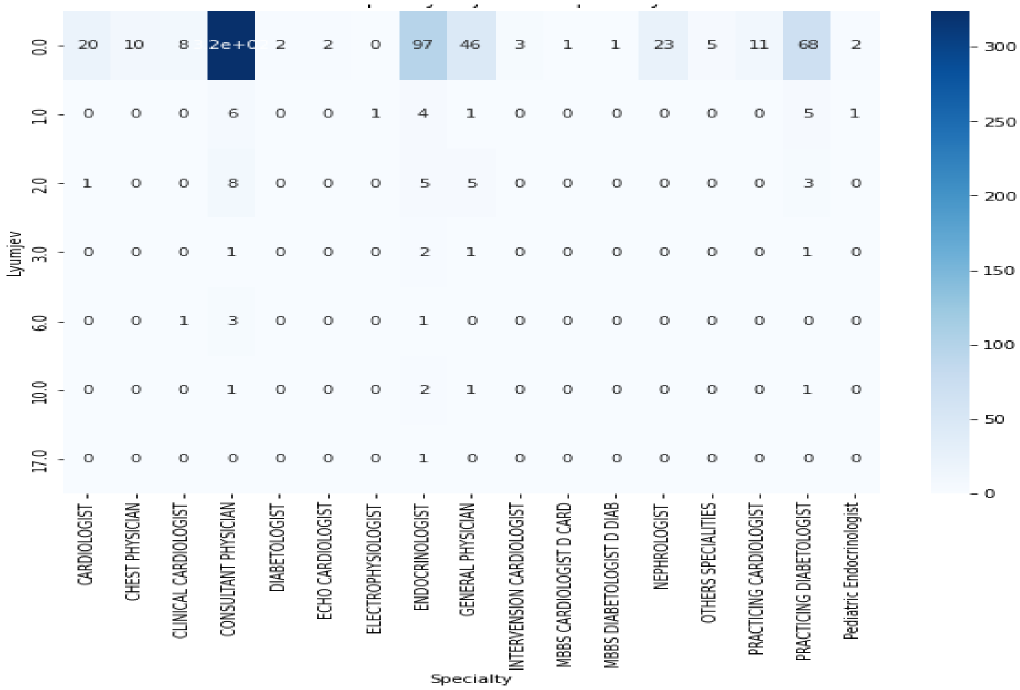

Figure 1: Heatmap for Classification Prescription Analysis

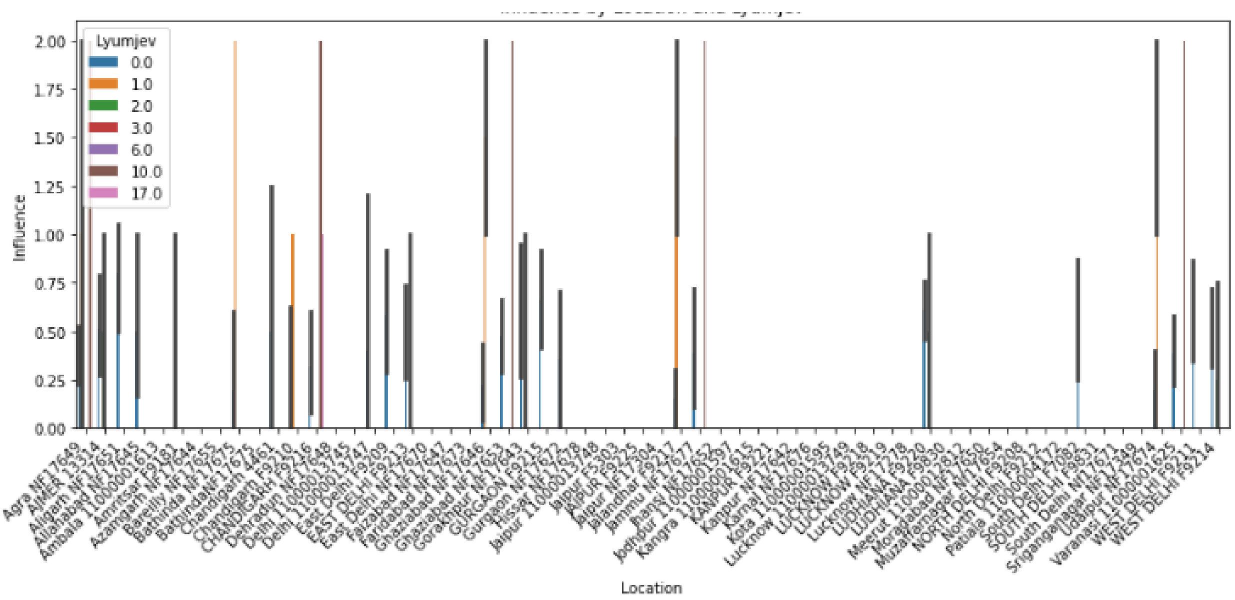

Figure2: Influence rate of Lyumjev in North Zone

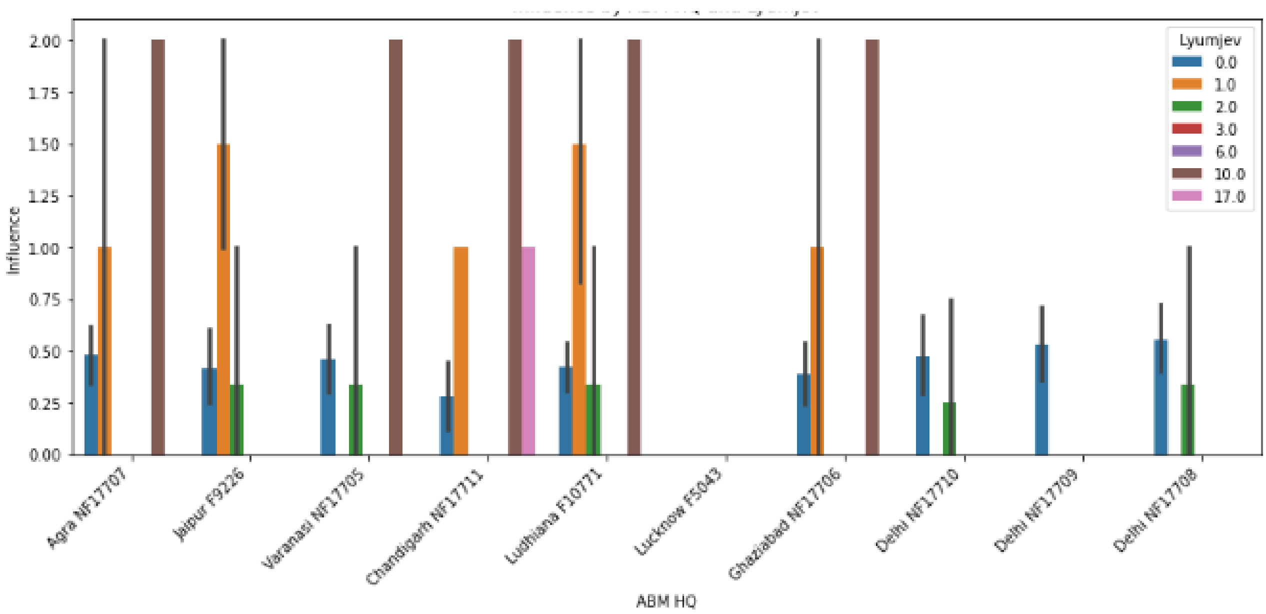

Figure3: Influence by ABM HQ and Lyumjev

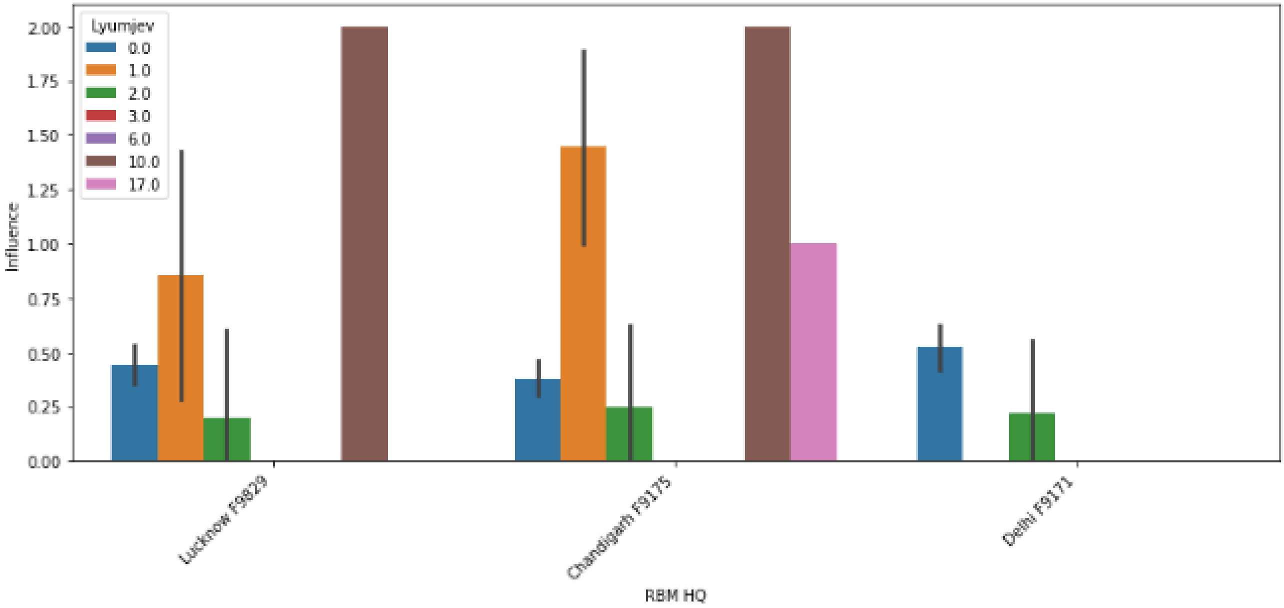

Figure4: Influence by RBM HQ and Lyumjev

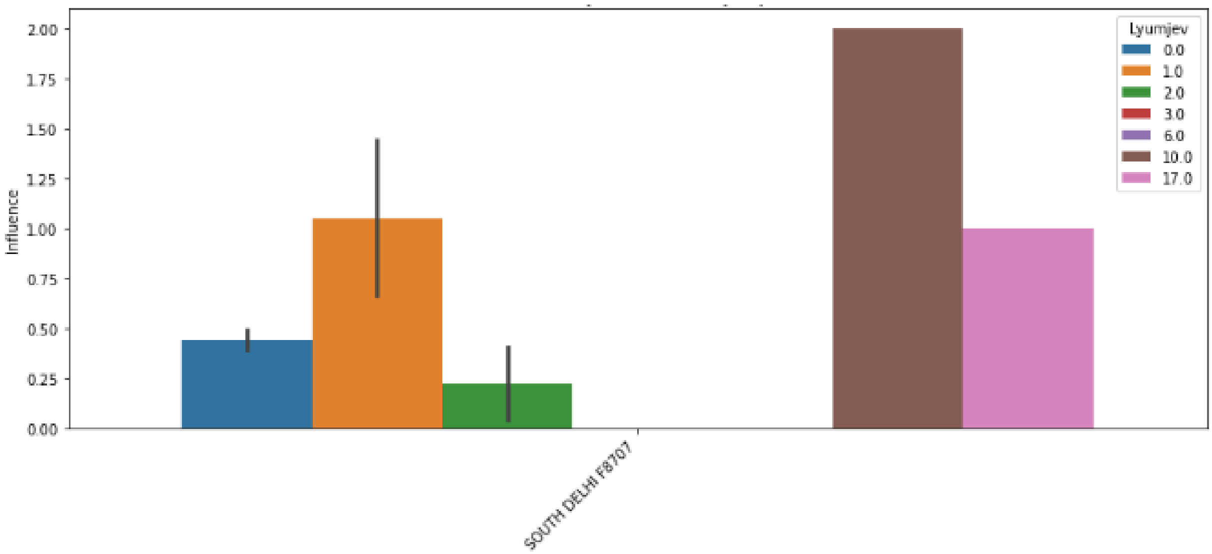

Figure5: Influence by ZBM HQ and Lyumjev

4. CONCLUSION

This research set out the examine rate of the influence rate of HCPs on the Lyumjev prescribing patterns in the north zone of India. It leverages the technique of machine learning models. The results provide insights into the patterns of the predictive model and can identify the influence rate, adoption, and prescribing trends of Lyumjev in a dynamic healthcare environment.

Machine learning models found major determinants such as physician specialty, patient demographics, and geographical considerations that influence HCP prescribing practices. Understanding these aspects allows pharmaceutical companies to better customize their marketing and engagement initiatives with HCPs, aligning more effectively with regional prescription trends and diabetes patients'changing requirements.

This study underscores the rising importance of machine learning in healthcare by providing actionable insights for optimizing medicine distribution and enhancing patient care [6]. The predictive models established here can be a useful tool for pharmaceutical firms, healthcare practitioners, and legislators who want to make data-driven decisions to increase the accessibility and use of innovative medicines like Lyumjev.

Future studies might improve these models by incorporating real-time data, broadening their geographic coverage, and applying them to additional therapeutic areas, resulting in a more detailed and complete knowledge of prescription trends in Indiaby Arora et al. [7]. It may also provide predictions about the doctor's adoption behavior, awareness, and HCP behavior toward the medicine. The models may be educated over the awareness of the drug and how the awareness of the drug can be part of the influence of the HCPs and can change the behavior patterns of the physicians prescribing pharmaceuticals.

5. REFERENCES

[1] E. Shairy and E. Gourav, "Unveiling Insights in the IoT Sector: Harnessing Power BI for Data Mining and

Visualization," *J. Adv. Res. Inf. Technol. Syst. Manage.*, vol. 8, no. 1, pp. 19–26, 2024. [Online]. Available: http://www.thejournalshouse.com/index.php/information-tech-systems-mngmt/article/view/1186

[2] M. Khazzaka, "Pharmaceutical marketing strategies' influence on physicians' prescribing pattern in Lebanon: ethics, gifts, and samples," *BMC Health Serv. Res.*, vol. 19, p. 80, 2019. [Online]. Available: https://doi.org/10.1186/s12913-019-3887-6

[3] B. L. Anderson, G. K. Silverman, G. F. Loewenstein, S. Zinberg, and J. Schulkin, "Factors associated with physicians' reliance on pharmaceutical sales representatives," *Acad. Med.*, vol. 84, no. 8, pp. 994–1002, Aug. 2009, doi: 10.1097/ACM.0b013e3181ace53a.

[4] D. K. Yimenu, C. A. Demeke, A. E. Kasahun, *et al*., "Health professional's exposure, attitude, and acceptance of drug promotion by industry representatives: A cross-sectional study in Ethiopia," *Sci. Prog.*, vol. 104, no. 2, 2021, doi: 10.1177/00368504211029435.

[5] A. R. Patwardhan, "Physicians-Pharmaceutical Sales Representatives Interactions and Conflict of Interest: Challenges and Solutions," *INQUIRY: J. Health Care Organ. Provis. Financ.*, vol. 53, 2016, doi: 10.1177/0046958016667597.

[6] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications, and Research Directions," *SN Comput. Sci.*, vol. 2, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.

[7] G. Arora, S. Kalra, and G. Kaur, "Introduction to Data Science and Analytics," in *Revolutionizing Data Science and Analytics for Industry Transformation*, P. Manuel, K. Qureshi, and T. Venkatachalam, Eds. IGI Global Scientific Publishing, 2025, pp. 1–28. doi: 10.4018/979-8-3693-7868-7.ch001.