Forecasting the Emergence of a Dominant Design by Classifying Product and Process Patents Using Machine Learning and Text Mining

Koji Masuda[1], Yoshinori Hayashi[2], Shigeyuki Haruyama[3]*
Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Ube, Japan[1, 2]
Graduate School of Innovation and Technology Management, Yamaguchi University, Ube, Japan[3]

***Abstract*—Forecasting the emergence of a dominant design in advance is important because the emergence of the dominant design can provide useful information about the external environment for the product launch. Although the emergence of the dominant design can only be determined as a result of the introduction of the product into the market, it may be possible to predict the emergence of the dominant design in advance by applying a solution based on patent analysis. In the newly proposed technique of separating patents, we can capture changes in the state of technological innovation and analyze the emergence of the dominant design, but there is a problem that it requires processing of large amounts of patent data, and that the processing involves subjective judgments by experts. This study focuses on analyzing technological innovation trends using an approach that separates product patents from process patents, investigates whether this approach can be applied to machine learning, and aims to develop a learning model that automatically classifies patents. We applied text mining to patent information to create structured data sets and compared nine different machine learning classification algorithms with and without dimensionality reduction. The approach was effectively applied to machine learning, and the Random Forest, AdaBoost and Support Vector Machine models achieved high classification performance of over 95%. By developing these learning models, it is possible to objectively forecast the emergence of a dominant design with high accuracy.**

Keywords*—*Dominant design; patent analysis; technological innovation; machine learning; text mining; classification

I. Introduction

A company's introduction of a product into a market can significantly change its competitive environment [1], while the external environment affects market entry [2], [3]. Thus, the timing of market entry is strategically important for companies [4]. Dominant design is defined as a design that has achieved market dominance [5], and some previous studies have discussed market entry timing in relation to dominant design. These studies point out that companies that enter the market when a dominant design is likely to emerge while timing their entry will win the market [6], and that entering the market just before the emergence of the dominant design is particularly advantageous and tends to have a low probability of failure [7]. However, the emergence of the dominant design is recognized as a result of a product's entry into the market and thus can only be known in retrospect [5], [8]. If the timing of market entry can be accurately predicted in advance, the probability of success can be increased by formulating and implementing a growth and technology strategies in accordance with that timing. Therefore, predicting the timing of the emergence of the dominant design is necessary.

Since the timing of the emergence of the dominant design is when the competitive advantage shifts from product innovation to process innovation [5], it is necessary to capture the change in the state of technological innovation in order to predict the emergence of the dominant design. Since patent information is important as an innovation indicator for companies [9] and is useful as an information source for predicting future products [1], patent analysis can be used to predict the state of technological innovation.

The problem of patent analysis, which is a complex and time-consuming process [10] and involves subjective and qualitative judgments of experts [11], [12], is well known. We propose a new technique for patent analysis that separates patents related to product innovation (product patents) from those related to process innovation (process patents) [13], and show that the timing of the emergence of the dominant design can be predicted using this technique by analyzing specific product case studies [14]. However, subjective processing by experts still remains, and there are concerns about the variability of the processing results. Patent analysis using automatic classification with machine learning allows for objective forecast of the emergence of a dominant design with high accuracy and stability. The increased efficiency provided by automatic classification contributes to reducing the activities and investments of companies for patent analysis.

In order to forecast the emergence of the dominant design, this study examines whether the idea can be applied to machine learning based on a technique for separating product patents from process patents and develops a learning model that automatically classifies patents into product patents and process patents. Specifically, we apply text mining to patent information, which is textual information, to extract features to be input to the modeling. We compare several classification algorithms for supervised learning and construct an appropriate learning model.

This paper is organized to provide a comprehensive understanding of analytical methods for automatically classifying patents into product and process patents using machine learning and text mining. Section I provides background and emphasizes the importance of predicting the emergence of the dominant design. Section II presents a literature review. Section III describes the methodology of the study, and Section IV presents the results and discussions. Section V presents the conclusions.

II. Literature Review

A. Dominant Design

There are previous studies that have analyzed the emergence of a dominant design based on patent information. In an analysis focusing on the number of patents per technology category, the dominant design is composed of technology categories with a large number of patents [15]. In an analysis focusing on the citation rate of patents, the dominant design exists when the ratio of patents citing the same patent in a patent class is 50% or more [16]. These analyze whether or not a dominant design emerges, but do not provide any information on the timing of the emergence of the dominant design.

The timing of the emergence of the dominant design is said to be the boundary between the fluid phase and the transition phase in “the dynamics of innovation” model [5]. Capturing changes in the state of technological innovation means that it may be possible to predict the timing of emergence by estimating the profiles of product and process innovation in the aforementioned model.

B. Classifying Product and Process Patents

In Japanese patent law, inventions are categorized into inventions of a product and inventions of a process, and inventions of a process are further categorized into inventions of a process that produces a product and inventions of a process that does not produce a product [17]. Patent laws in Europe and the United States categorize inventions in almost the same way [18], [19]. Product inventions are inventions relating to the product itself. Process inventions, on the other hand, refer to inventions relating to a process for manufacturing or producing a product, inventions relating to a process for improving or enhancing the characteristics of a product, and inventions relating to a process for expressing the function of a product, based on the content of the invention.

The patents related to product innovation and process innovation in the "dynamics of innovation" model are called product patents and process patents, and the two types of patents are shown in Table I, which maps them to the various inventions mentioned above.

Previous studies on the classification of product and process patents propose methods for experts and specialists to judge their classification, and they focus on the description of the F-term, which is a Japanese patent classification code [20], or on the title of the invention [14]. In both cases, the large amount of patent data has to be processed subjectively by experts, and there are concerns about the stability and efficiency of the processing results.

C. Machine Learning for Patent Analysis

Previous studies point out that patent analysis requires very large data sets and expertise, and that manual, subjective processing is time-consuming and costly [21], [22], [23], thus automation using machine learning is eagerly awaited. The focus of patent analysis is on extracting specific technology information and investigating technology trends [24], and the analysis of "technology" is the main objective. For example, the following are examples of patent analysis using machine learning. In terms of technology information extraction, there is the extraction of vacant technology [25], the identification of emerging technologies [26], and the extraction of differences in technologies of competing companies [27]. In addition, for technology trend studies, there are the future technology trends in a certain industry [28], the trajectory of technology development from the present to the future [29], the current and future technology impact in a certain technology field [30], and the prediction of technology convergence in a certain industry or technology field [31].

TABLE I. Classification of Product and Process Patents

<table>
<tr><th colspan="3">Categories of Invention</th><th>Contents of Invention</th><th>Classification of Patent</th></tr>
<tr><td rowspan="4">Inventions</td><td rowspan="3">Inventions of a process</td><td>Inventions of a production process</td><td>Inventions that manufacture or produce a product</td><td rowspan="2">Process patents</td></tr>
<tr><td rowspan="2">Inventions of a non-production process</td><td>Inventions that improve or enhance the characteristics of a product</td></tr>
<tr><td>Inventions that express the function of a product</td><td rowspan="2">Product patents</td></tr>
<tr><td colspan="2">Inventions of a product</td><td>Inventions relating to the product itself</td></tr>
</table>

We study the use of patent information not to analyze technologies for R&D, but to analyze innovations as value creation for customers, markets, and society [14]. In conventional patent analysis using machine learning, the main target of analysis is the investigation of technology trends, while few research reports are known to focus on the analysis of innovation. We focus on technological innovation in the analysis of patents using machine learning, especially in the investigation of innovation trends as shown in "the dynamics of innovation" model.

In the next section, we describe a patent analysis method that focuses on " title of the invention" as patent information, and automatically classifies patents into product patents and process patents by using machine learning and text mining.

III. Method

This study followed the process model developed by the CRoss Industry Standard Process for Data Mining (CRISP-DM) project [32], which was a de facto standard process model for data mining projects that can be applied independently of industries and research domains [33]. Table II shows an overview of the individual phases of CRISP-DM, which consists of six phases, as well as the general tasks [34].

The following subsections described the methodology of this study for each phase.

TABLE II. PROCESS MODEL OF CRISP-DM

Phase	Outline and Generic Task
Business understanding	The business understanding phase focuses on understanding the objectives and requirements of the project from a business perspective, then developing data mining objectives and creating a plan, including an initial evaluation of tools and techniques, to achieve the objectives.
Data understanding	The data understanding phase begins with collecting the data to be used in the analysis, organizing the characteristics of the data to become familiar with the data, and performing simple tabulations. Activities proceed to understanding the meaning of the data and checking the quality of the data.
Data preparation	The data preparation phase includes all activities to prepare the final data set (the data supplied to the modeling tool) from the initial data. These activities include data selection, data cleaning, data construction, data integration, and data transformation.
Modeling	The modeling phase involves selecting and applying different modeling techniques and adjusting their parameters to optimal values. In general, there are several techniques for the same type of data mining problem. In the case of supervised learning, the data sets are usually divided into training and test data set, the model is built on the training data set, and its quality is estimated on the test data set. Metrics to evaluate the quality and validity of the model are generated before the model is built.
Evaluation	During the evaluation phase, it is important to review the steps taken to ensure that the model adequately achieves the business objectives. A more detailed review of data mining is appropriate to determine if any tasks have been overlooked.
Deployment	During the deployment phase, the process of building the model is documented, the entire project is reviewed, and a final report is compiled for future use.

A. Business Understanding

The objectives of the data analysis project were understood, and the resources and constraints for implementation were identified. Next, the data mining objectives were determined from a technical perspective and an action plan was developed. As a source of patent information, registered patents on projector products that predict and validate the emergence of the dominant design were selected [14]. An initial evaluation of tools and techniques was performed during this phase.

B. Data Understanding

As shown in Table I, inventions are classified into inventions of a product and inventions of a process. This classification can be easily made by paying attention to the "title of the invention" in the patent specification. In other words, it can be determined whether the keyword "process" is included in the "title of the invention" or not. Based on this understanding of the data, the "title of the invention" data of each patent was collected as the patent information to be used in the analysis.

Inventions of a product and inventions of a process were simply tabulated. Data quality was checked for completeness and missing values.

All inventions of a product belong to product patents. On the other hand, inventions of a process belong either to product patents or to process patents. Therefore, it is necessary to determine from the content of the "title of the invention" whether the patent containing the invention is a product patent or a process patent.

C. Data Preparation

For the inventions of a process, the following two structured data sets were constructed to prepare a final data set from the collected " title of the invention" data. They were combined into the final data set.

One was a high-dimensional structured data set created by tokenization, data cleaning and feature extraction with TF-IDF (Term Frequency – Inverted Document Frequency) using text mining techniques on the unstructured text data of the "title of the invention". TF-IDF is a feature that assigns a lower weight to words that appear in more documents relative to the frequency of occurrence of the word. TF_{ij} is denoted by tf_{ij}, the frequency of the word w_j in document d_i (1), and IDF_j is denoted by Eq. (2), where N is the total number of documents and df_j is the number of documents containing the word w_j [35]. $TF\text{-}IDF_{ij}$ is denoted by Eq. (3), where the document refers the "title of the invention".

$$TF_{ij} = tf_{ij} \tag{1}$$

$$IDF_j = log(1+N/df_j) \tag{2}$$

$$TF\text{-}IDF_{ij} = TF_{ij} * IDF_j = tf_{ij} * log(1+N/df_j) \tag{3}$$

The other was a single row of structured data set formed by labeling whether the patent containing the inventions of a process method was a product patent or a process patent. The labeling was performed by engineers and experts familiar with the technology.

D. Modeling

We conducted modeling to classify patents containing inventions of a process into product patents and process patents using a machine learning algorithm. We investigated the well-known supervised learning classifiers: Decision Trees (DT), Linear Discriminant (LD), Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Random Forest (RF), AdaBoost (AB), and Neural Networks (NN) [36], [37], [38], [39]. The data set created in the previous subsection was used as input, and the hyperparameters were tuned for each classification model using Bayesian optimization.

Because dimensionality reduction has the potential to improve model performance, all models were run with and without dimensionality reduction using Principal Component Analysis (PCA) on the input data set.

The ratio of training and test data sets was set to 80% and 20%. To avoid overfitting, a five-fold cross-validation was used for training. That is, the 80% training data set is divided into 64% for training and 16% for validation. The Mean Accuracy of the cross-validation on the training data set was calculated as a metrics of the quality of the classification model. In addition, we calculated Accuracy using the test data set, Recall, which indicates how well the model reproduces actual results, Precision, which indicates how well the model corrects predicted results, and F1-Score, which is the harmonic mean of Precision and Recall with a trade-off relationship. The confusion matrix shown in Table III and the following Eq. (4), (5), (6) and (7) were used in these calculations.

TABLE III. CONFUSION MATRIX

		Prediction	
		Positive	Negative
Actual	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Positive (FP)	True Negative (FP)

$$Accuracy = (TP + TN) / (TP + FN + FP + TN) \quad (4)$$

$$Precision = TP / (TP + FP) \quad (5)$$

$$Recall = TP / (TP + FN) \quad (6)$$

$$F1\text{-}Score = 2 * TP / (2 * TP + FP + FN) \quad (7)$$

The classification models with good values for these metrics were selected.

E. Evaluation

To confirm that the business objective of forecasting the emergence of a dominant design was feasible, each step was reviewed and confirmed.

F. Deployment

The entire project, including the procedures for data preparation by text mining and modeling by machine learning, was summarized in this paper.

IV. RESULTS AND DISCUSSIONS

The results were presented in the order of the phases outlined in the previous section, followed by some discussion.

A. Business Understanding

The objective of data mining is the automatic classification of product patents and process patents, and in particular the classification of " inventions of a process " into product patents and process patents. For the initial evaluation of the tools and techniques, we conducted a preliminary experiment on 1,000 registered patents for projectors, which is a simplified version of a planned main experiment and confirmed that the planned experiment was feasible. In the preliminary experiment, we went through the procedures of data preparation, modeling, and evaluation, and found that it was likely to provide the desired accuracy, thus we decided to proceed with the main experiment. We used MATLAB R2023b version, Statistics and Machine Learning Toolbox, and Text Analytics Toolbox from Mathworks as the tools to perform text mining and machine learning.

B. Data Understanding

Registered patents were extracted from the Japan Patent Office (JPO) database using the search conditions of patent classification code and period. For the patent classification codes, we used theme codes that are unique to Japan. Theme codes are organized by technical groupings and can be represented almost equivalently by a bundle of multiple IPCs. The number of registered patents extracted under the conditions shown in Table IV was 11,318. Based on a simple aggregation by the presence or absence of the keyword "process" in the "title of the invention", 8,932 patents were classified as inventions of a product, and 2,386 patents were classified as inventions of a process.

TABLE IV. SEARCH CONDITIONS

Item	Query
Database	Japan Patent Office
Patent classification code (Theme code)	2K103 or 2K203
Period	1/1/1981 – 12/31/2020
Search date	10/30/2023

The “title of the invention,” which includes both inventions of a product and inventions of a process, was positioned as inventions of a process. This is because inventions of a process are classified as product patents and process patents in the next phase of modeling.

C. Data Preparation

For the 2,386 unstructured text data of “title of the invention,” we performed cleaning and computed TF-IDF to create a structured data matrix of numerical variables with 2,386*714 dimensions. After tokenization, the cleaning process included stemming, erasing punctuation, removing stop words, removing a single character, and standardizing synonyms.

Labeling was performed by experts to create a 2,386*1 dimensional structured data matrix with process patents as “A” and product patents as “B.” By merging the two structured data sets, a 2,386*715-dimensional matrix was created as the final data set for the modeling.

D. Modeling

Table V shows the mean accuracy of each model on the training data set for each of the nine classifiers. To check the effect of dimensionality reduction, PCA was performed on the input data set to achieve a cumulative contribution rate of at least 95%, and the dimensionality was reduced from 714 dimensions to 343 dimensions. All models were run without dimensionality reduction (without PCA) and with dimensionality reduction (with PCA).

Only NB and kNN had mean accuracy below 90%, while the rest of the models exceeded 90%. In particular, the AB model achieved good mean accuracy of over 95%.

Table VI shows the calculation results for each metric on the test data set. For each classification algorithm, the model with the higher mean accuracy was selected with and without PCA. In the case of with PCA, "with PCA" was added to the name of the classifier.

TABLE V. COMPARISON BETWEEN WITHOUT PCA AND WITH PCA (TRAINING DATASET)

Classification Algorithm (Classifier)	Mean Accuracy on the Training Data Set	
	Without PCA	**With PCA**
DT	94.1%	85.8%
LD	85.7%	93.8%
LR	89.3%	91.0%
NB	70.7%	84.7%
SVM	94.9%	94.1%
kNN	89.1%	89.6%
RF	94.7%	89.9%
AB	95.1%	92.1%
NN	94.0%	94.1%

TABLE VI. COMPARISON BETWEEN WITHOUT PCA AND WITH PCA (TEST DATASET)

Classification Algorithm (Classifier)	On the Test Data Set			
	Accuracy	**Precision**	**Recall**	**F1-score**
DT	94.6%	96.3%	95.7%	96.0%
LD with PCA	93.1%	96.2%	93.5%	94.8%
LR with PCA	93.1%	95.3%	94.4%	94.9%
NB with PCA	81.6%	82.9%	91.6%	87.1%
SVM	95.6%	95.8%	97.8%	96.8%
kNN with PCA	91.0%	92.4%	94.4%	93.4%
RF	95.6%	97.8%	95.7%	96.7%
AB	95.2%	96.9%	96.0%	96.4%
NN with PCA	94.3%	94.6%	97.2%	95.9%

Accuracy was highest for SVM and RF, followed by AB at more than 95%. Precision was highest for RF, and AB, DT, LD with PCA, SVM, and LR with PCA exceeded 95%. Recall was highest for SVM, followed by NN with PCA, AB, RF, and DT over 95%. The F1-Score, the harmonic mean of Precision and Recall, was also highest for SVM, followed by RF, AB, DT, and NN with PCA exceeding 95%.

Tables VII, VIII, and IX show the confusion matrices on the test data set for the three models SVM, RF, and AB, which performed well above 95% on all four metrics.

The high performance of several models in patent classification in this study suggested that the "title of the invention" was appropriate as patent information data, that the data preprocessing was effective, and that the idea of separating product and process patents was applicable to machine learning.

In this experiment, which combined nine classification algorithms with and without PCA, the prediction model using SVM, RF and AB algorithms achieved higher performance.

TABLE VII. CONFUSION MATRIX OF SVM

		Prediction	
		A	B
Actual	A	316	7
	B	14	140

TABLE VIII. CONFUSION MATRIX OF RF

		Prediction	
		A	B
Actual	A	309	14
	B	7	147

TABLE IX. CONFUSION MATRIX OF AB

		Prediction	
		A	B
Actual	A	310	13
	B	10	144

E. Evaluation

In order to forecast the emergence of a dominant design, which is the objective of the business, it was important to capture changes in the state of innovation. The changes were indicated by the trends of product patents and process patents according to the “the dynamics of innovation” model. Based on Table I, we categorized the patents to be analyzed into product patents and process patents. Since inventions of a product can be easily identified from the “title of the invention,” we focused on classifying inventions of a process into product patents and process patents. The data preparation and modeling resulted in several prediction models with high classification performance in terms of the overall model correctness rate and the F1-Score, which is a balance between actual and predicted results.

Since the trends of product innovation and process innovation are visualized according to the classification results of the prediction model, and the emergence of a dominant design is predicted, we considered precision to be particularly important among the four metrics for this business objective. Therefore, the prediction model with the highest precision performance was preferred. Table VI shows that the precision performance of the RF model is 97.8%, and the predicted trends of product and process patents are almost the same as their actual trends. The above review confirmed that no tasks were missed in the steps performed and that the business objective was properly achieved. It also demonstrated that the automatic classification by machine learning worked effectively.

F. Deployment

In this project, the business objective was to predict the emergence of a dominant design, and the data mining goal for this purpose was to automatically classify "inventions of a process" into product patents and process patents. Through data understanding, data preparation, modeling, and evaluation, the validity of the data we focused on and the predictive models that achieved high performance were confirmed, and thus the project was completed. We summarized the data mining process and results according to the CRISP-DM process model in this paper.

G. Discussions

The effect of dimensionality reduction on the classification algorithm was discussed by comparing the models with and without PCA. Table V shows that the five models with a higher mean accuracy with PCA than without PCA were LD, LR, NB, kNN, and NN. According to the idea that machine learning models can be divided into three models: geometric, probabilistic, and logical models [40], these five models were included in the geometric and probabilistic models. The models with a difference of less than 1% between those with and without PCA were SVM, kNN, and NN, all of which were geometric models. On the other hand, four models, DT, SVM, RF, and AB, had a higher mean accuracy without PCA than with PCA. Since RF and AB are ensemble learning with tree models, these three models including DT are considered to be logical models. These results suggested that dimensionality reduction may be effective in improving the performance of geometric and probabilistic models in this experiment.

It is known that SVM and ensemble learning, such as RF and AB, tend to show relatively high performance compared to other algorithms, and this study was consistent with this finding, as well as previous studies comparing multiple algorithms [36],[37].

Although this study achieved good results in classification performance, some limitations need to be considered. Instead of classifying product inventions and process inventions directly from the "title of the invention," this study focused on separating "inventions of a product" and "inventions of a process" from the "title of the invention" by a simple procedure first, and then classifying product inventions and process inventions from "inventions of a process." We used TF-IDF and five-fold cross-validation for feature extraction in data preparation and data partitioning in modeling, respectively, but other techniques could be considered to further improve classification performance.

V. Conclusions

In order to forecast the emergence of dominant designs, this study investigated an automatic classification method for product and process patents according to the CRISP-DM process model applied to data mining projects. We focused on "title of the invention" as patent information, extracted TF-IDF features by text mining, and evaluated nine classification algorithms with and without PCA by machine learning. As a result, the prediction model using the RF, AB, and SVM algorithms achieved over 95% performance in all four metrics: accuracy, precision, recall, and F1-Score. In the classification of product patents and process patents, it was shown that the "title of the invention" was appropriate as patent information data, that data preprocessing was effective, and that the idea of a technique for separating product patents from process patents was applicable to machine learning.

By using patent analysis, which uses machine learning and text mining to capture changes in product innovation and process innovation, that is, changes in the state of technological innovation, it is possible to objectively forecast the emergence of a dominant design with high accuracy. Therefore, it can be a useful piece of information about the external environment for companies to formulate and implement growth and technology strategies.

Increased efficiency in analyzing trends in technological innovation can lead to a reduction in the activities and investments of companies. In addition, the resources generated by the reduction are expected to make a new contribution.

References

[1] J. M. Gerken, M. G. Moehrle, and L. Walter, "One year ahead! Investigating the time lag between patent publication and market launch: Insights from a longitudinal study in the automotive industry," R&D Management, vol. 45, no.3, pp. 287-303, 2015.

[2] B. L. Bayus, "Speed-to-market and new product performance trade-offs," Journal of Product Innovation Management, vol. 14, no. 6, pp. 485-497, 1997.

[3] D. Morschett, H Schramm-Klein, and B. Swoboda, "Decades of research on market entry modes: What do we really know about external antecedents of entry mode choice?," Journal of International Management, vol. 16, no. 1, pp. 60-77, 2010.

[4] F. F. Suarez, S. Grodal, and A. Gotsopoulos, "Perfect timing? Dominant category, dominant design, and the window of opportunity for firm entry," Stragetic Management Journal, vol. 36, no. 3, pp. 437-448, 2015.

[5] J. M. Utterback, "Mastering the dynamics of innovation," Boston, MA, USA: Harvard Business Review Press, 1994.

[6] C. C. Markides and P. A. Geroski, "Fast second: How smart companies bypass radical innovation to enter and dominate new markets," San Francisco, CA, USA: Jossey-Bass, 2004.

[7] C. M. Christensen, F. F. Suarez, and J. M. Utterback, "Strategies for survival in fast-changing industries," Management Science, vol. 44, no. 12-part-2, pp. 207-220, Dec. 1998.

[8] P. Anderson and M. L. Tushman, "Technological discontinuities and dominant designs: A cyclical model of technological change," Administrative Science Quaterly, vol. 35, no. 4, pp. 604-633, Dec. 1990.

[9] S. Rocheska, D. Nikoloski, M. Angeleski, and G. Mancheski, "Factors affecting innovation and patent propensity of SMEs: Evidence from Macedonia," TEM Journal, vol. 6, no. 2, pp. 407-415, May 2017.

[10] K. Masuda and S. Haruyama, "Forecasting technology trends based on separation of product inventions and process inventions: The technology S-curve," IOP Conference Series: Materials Science and Engineering, 1034, 012123, 2021.

[11] K. Masuda and S. Haruyama, "Forecasting the timing of the emergence of a dominant design: The case of the projectors," International Journal of Technology, vol. 15, no. 1, pp. 138-153, 2024.

[12] N. Clymer and S. Asaba, "A new approach for understanding dominant design: The case of the ink-jet printer," Journal of Engineering and Technology Management, vol. 25, no. 3, pp. 137-156, 2008.

[13] A. Brem, P. A. Nylund, and G. Schuster, "Innovation and de facto strandardization: The influence of dominant design on innovative performance, radical innovation, and process innovation," Technovation, vol. 50-51, pp. 79-88, 2016.

[14] The Japan Patent Office, "Introduction to the intellectual property act," Accessed: Jul. 1, 2024. [Online]. Available: https://www.jpo.go.jp/news/kokusai/developing/training/textbook/document/index/Introduction_to_The_Intellectual_Property_Act.pdf

[15] The European Patent Office, "Guidelines for examination in the European Patent Office," Accessed: Jul. 1, 2024. [Online]. Available: https://www.epo.org/en/legal/guidelines-epc

[16] The United States Patent and Trademark Office, "Manual of patent examining procedure (MPEP)," Accessed: Jul. 1, 2024. [Online]. Available: https://www.uspto.gov/web/offices/pac/mpep/index.html

[17] Y. Ishii, K. Kaminishi, and S. Haruyama, "A study of identifying trends in projector using F-term codes from Japanese patent applications," International Journal of Integrated Engineering," vol. 13, no. 7, pp. 324-332, 2021.

[18] X. Zhang, "Interactive patent classification based on multi-classifier fusion and active learning," Neurocomputing, vol. 127, pp. 200-205, 2014.

[19] J.-L. Wu, P.-C. Chang, C.-C. Tsao, and C.-Y. Fan, "A patent quality analysis and classification system using self-organizing maps with support vector machine," Applied Soft Computing, vol. 41, pp. 305-316, 2016.

[20] A. J. C. Trappey, C. V. Trappey, J.-L. Wu, and J. W. C. Wang, "Intelligent compilation of patent summaries using machine learning and natural language processing techniques," Advanced Engineering Infomatics, vol. 43, 101027, 2020.

[21] J. Choi, D. Jang, S. Jun, and S. Park, "A predictive model of technology transfer using patent analysis," Sustainability, vol. 7, 16175, 2015.

[22] S. Jun and S.-S. Park, "Examining technological innovation of Apple using patent analysis," Industrial Management & Data Systems, vol. 113, No. 6, pp. 890-907, 2013.

[23] C. Lee, O. Kwon, M. Kim, and D. Kwon, "Early identification of emerging technologies: A machine learning approach using multiple patent indicators," Technological Forecasting & Social Change, vol. 127, pp. 291-303, 2018.

[24] S. Jun and S.-S.- Park, "Examining technological competition between BMW and Hyundai in the Korean car market," Technology Analysis & Strategic Management, vol. 28, no. 2, pp. 156-175, 2016.

[25] A. suominen, H. Toivanen, and M. Seppanen, "Firms' knowledge profiles: Mapping patent data with unsupervised learning," Technological Forecasting & Social Change, vol. 115, pp. 131-142, 2017.

[26] L. J. Aaldering and C. H. Song, "Tracing the technological development trajectory in post-lithium-ion battery technologies: A patent-based approach," Jounal of Cleaner Production, vol. 241, 118343, 2019.

[27] D. Thorleuchter, D. Van den Poel, and A Prinzie, "A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies," Technological Forecasting & Social Change, vol. 77, pp. 1037-1050, 2010.

[28] T. S. Kim and S. Y. Sohn, "Machine-learning-based deep semantic analysis approach for forecasting new technology convergence," Technological Forecasting & Social Change, vol. 157, 120095, 2020.

[29] R. Wirth and J. Hipp, "CRISP-DM: Towords a standard process model for data mining," in Proceedings of the Fourth International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Manchester, UK, 2000, pp. 29-40.

[30] C. Schroer, F. Kruse, and J. M. Gomez, "A systematic literature review on applying CRISP-DM process model," Procedia Computer Science, vol. 181, pp. 526-534, 2021.

[31] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. CRISP-DM 1.0: Step-by-step data mining guide. (2000).

[32] MathWorks. Help Center. Accessed: Jul. 1, 2024. [Online]. Available: https://www.mathworks.com/help/index.html

[33] Y. Mahmud, N. S. Shaeeali, and S. Mutalib, "Comparison of machine learning algorithms for sentiment classification on fake news detection," International Journal of Advanced Computer Science and Applications, vol. 12, no. 10, pp. 658-665, 2021.

[34] M. Sivamanikanta and N. Ravinder, "Machine learning-driven integration of genetic and textual data for enhanced genetic variation classification," International Journal of Advanced Computer Science and Applications, vol. 15, no. 1, pp. 252-262, 2024.

[35] M. Chistol and M. Danubianu, "Automated detection of autism spectrum disorder symptoms using text mining and machine learning for early diagnosis," International Journal of Advanced Computer Science and Applications, vol. 15, no. 2, pp. 610-617, 2024.

[36] S. Fadili, M. Ertel, A. Mengad, and S. Amali, "Predicting optimal learning approaches for nursing students in Morocco," International Journal of Advanced Computer Science and Applications, vol. 15, no. 4, pp. 94-102, 2024.

[37] P. Flach, "Machine learning: The art and science of algorithms that make sense of data," New York, NY, USA: Cambridge University Press, 2012.

[38] K. OuYang and C. S. Weng, "A new comprehensive patent analysis approach for new product design in mechanical engineering," Technological Forecasting & Social Change, vol. 78, pp. 1183-1199, 2011.

[39] G. Kim and J. Bae, "A novel approach to forecast promising technology through patent analysis," Technological Forecasting & Social Change, vol. 117, pp. 228-237, 2017.

[40] J. Lee, N. Ko, J. Yoon, and C. Son, "An approach for discovering firm-specific technology opportunities: Application of link prediction to F-term networks," Technological Forecasting & Social Change, vol. 168, 120746, 2021.