MANTECH Publications

Journal of Data Mining and Knowledge Engineering
Volume 1, Issue 2, July-December 2016

Explainable Pattern Mining and Interpretable Models

Avinash K Pandey
Associate Professor
Department of Information Technology
Shanti College of Engineering, India
Email: *Avinashk055pandey@gmail.com*

Abstract

Pattern mining and machine learning models are widely used in diverse domains such as healthcare, finance, and cybersecurity to extract meaningful insights from complex datasets. However, the increasing complexity of models has raised concerns about their interpretability and transparency. Explainable pattern mining (XPM) and interpretable models provide methods to not only uncover patterns but also make the results understandable for humans. This review paper explores the current research trends in explainable pattern mining, examines different interpretable modeling techniques, and discusses their applications, benefits, and limitations. The paper also highlights challenges and future research directions in building models that balance predictive performance with explainability.

Keywords: *Explainable AI, Pattern Mining, Interpretable Models, Transparent Machine Learning, Knowledge Discovery, Model Explainability*

INTRODUCTION

The rise of machine learning and data-driven decision-making has transformed many industries, enabling automated predictions, classifications, and recommendations. Traditional pattern mining techniques, including association rule mining, sequential pattern mining, and frequent itemset mining, are often used to discover hidden structures in data. However, as predictive models become more complex (e.g., deep learning models), understanding why a model makes certain predictions has become a critical issue.

Explainable pattern mining focuses on extracting patterns from data that are comprehensible

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

and actionable by humans. Similarly, interpretable models are designed so that their internal mechanics can be easily understood without the need for specialized tools. Together, XPM and interpretable models aim to bridge the gap between predictive performance and human comprehension.

The primary objectives of this paper are:

- To review state-of-the-art methods in explainable pattern mining.
- To explore interpretable modeling approaches across various domains.
- To discuss applications, challenges, and future directions in the field.

BACKGROUND AND FUNDAMENTALS

Explainable pattern mining and interpretable models build upon foundational concepts in pattern discovery, machine learning, and explainable AI. Before discussing advanced methods, it is important to understand the underlying principles and techniques that enable extraction of meaningful patterns and transparent decision-making.

1. Pattern Mining

Pattern mining is the process of discovering **frequent, interesting, or useful structures** within large datasets. These structures, commonly referred to as patterns, reveal relationships or regularities among data items. Pattern mining is widely used in diverse domains such as retail, bioinformatics, social networks, and sensor networks. The primary goal is to uncover patterns that are **informative and actionable** for decision-makers.

Some of the key pattern mining tasks are:

- **Frequent Itemset Mining (FIM):**
  Frequent itemset mining aims to find groups of items that appear together frequently in transactional data. For example, in a supermarket dataset, the combination {Bread, Milk} may appear together in 25% of all transactions. Algorithms such as Apriori and FP-Growth are commonly used to discover frequent itemsets. FIM forms the basis of association rule mining, which can provide human-interpretable rules for marketing and recommendation systems.

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

- **Sequential Pattern Mining (SPM):**
  Sequential pattern mining focuses on identifying **ordered sequences of events** that occur frequently. For instance, an online retailer may find that customers often buy a smartphone, then a phone case, and later a screen protector. SPM can reveal temporal trends or user behaviors, which are critical in applications like clickstream analysis, recommendation engines, and medical event prediction. Algorithms like PrefixSpan and SPADE are widely used for this purpose.

- **Graph Pattern Mining:**
  Graph pattern mining involves discovering recurring subgraphs or motifs in networked data. Examples include social networks, protein interaction networks, and communication networks. Identifying frequent subgraphs can help detect communities, functional modules, or anomalous patterns in network structures. Techniques such as gSpan and GraphMiner are frequently employed to mine graph-based patterns.

Although pattern mining methods are powerful, **high-dimensional and unstructured datasets** often produce patterns that are difficult to interpret. Without careful processing or simplification, the discovered patterns may be too numerous or complex for practical decision-making. Explainable pattern mining addresses this by focusing on **patterns that are not only frequent but also comprehensible** to humans.

2. Interpretable Models

Interpretable models are **machine learning models designed to be understandable and transparent**, allowing humans to reason about how input features influence predictions. In contrast to black-box models, which may achieve higher accuracy but provide little insight into their internal mechanisms, interpretable models prioritize **explainability** alongside predictive performance.

Some common types of interpretable models include:

1. **Linear Models:**

Linear regression and logistic regression are classic interpretable models. In these models, the effect of each feature is represented by a coefficient, which indicates the **strength and direction** of its impact on the prediction. For example, in a medical dataset, the coefficient of

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

"age" in a logistic regression predicting heart disease risk would quantify how much increasing age increases risk. The simplicity of linear models makes them **easy to explain**, but they may fail to capture non-linear or complex interactions between features.

2. **Decision Trees:**

   Decision trees represent decisions as a hierarchical structure of nodes and branches. Each internal node corresponds to a feature-based decision, and each leaf node represents a prediction outcome. Decision trees are visually intuitive and provide **clear, step-by-step explanations** for how a prediction is made. However, very deep trees can become complex and less interpretable.

3. **Rule-Based Models:**

   Rule-based models express knowledge as a set of **IF-THEN rules**, which are straightforward for humans to understand. For instance:

IF Age > 50 AND Cholesterol > 240 THEN Risk = High

Rule-based models are widely used in expert systems, healthcare, and fraud detection due to their **direct interpretability and ease of reasoning**.

Complex models, such as **deep neural networks or ensemble methods**, often provide higher predictive performance but behave as black boxes. Their internal computations are difficult for humans to interpret directly. To overcome this, post-hoc explanation techniques are often required to make these models interpretable.

3. Explainability in AI

Explainable Artificial Intelligence (XAI) is a field that aims to **make machine learning models transparent and understandable** to humans. The motivation for XAI arises from the need to **trust and verify** model predictions, especially in critical domains like healthcare, finance, and law.

XAI techniques are broadly divided into:

1. **Intrinsic Explainability:**

These are models that are **interpretable by design**, such as linear models, decision trees, and

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

rule-based classifiers. Since the model structure itself is understandable, users can directly reason about predictions. Intrinsic models are suitable when interpretability is prioritized over raw accuracy.

2. **Post-hoc Explainability:**

These methods provide explanations for **already trained black-box models**. Examples include:

- **LIME (Local Interpretable Model-Agnostic Explanations):** Explains individual predictions by approximating the complex model with a simpler interpretable model locally around a prediction.
- **SHAP (SHapley Additive ExPlanations):** Uses game-theoretic principles to assign contribution scores to features, indicating their impact on the model's output.
- **Counterfactual Explanations:** Identify minimal changes to input features that would alter the prediction outcome, helping understand model behavior in "what-if" scenarios.

Explainable pattern mining **integrates traditional pattern mining with XAI techniques**, ensuring that the patterns discovered are not only statistically significant but also **human-interpretable**. For example, combining sequential pattern mining with SHAP explanations can help users understand which sequences of events most strongly influence predictions in temporal data.

In summary, the **synergy between pattern mining, interpretable models, and explainable AI** is crucial for building systems that provide actionable insights while maintaining transparency and trustworthiness.

EXPLAINABLE PATTERN MINING TECHNIQUES

Explainable pattern mining (XPM) aims to discover patterns from data that are **not only statistically significant but also understandable and actionable** by humans. Depending on the methodology and output format, XPM techniques can be broadly categorized into three groups: **rule-based, model-based, and hybrid approaches**. Each category offers different strengths and trade-offs in terms of interpretability, scalability, and predictive utility.

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Rule-Based Pattern Mining

Rule-based methods focus on generating **human-readable patterns** that can be easil interpreted. These approaches are especially suitable when the goal is to provide **explicit knowledge** about relationships in the dataset. They are widely used in domains such as retail, healthcare, finance, and cybersecurity where transparency and explanation are crucial.

Some of the most common rule-based techniques include:

1. Association Rules

Association rule mining is one of the oldest and most widely used methods for pattern discovery. It extracts **IF-THEN rules** from transactional datasets, highlighting frequent co-occurrences of items or events.

Definition:

- **Support:** The proportion of transactions that contain a particular itemset.
- **Confidence:** The conditional probability that the rule's consequent occurs given the antecedent.
- **Lift:** Measures how much more often the antecedent and consequent occur together than expected if they were independent.

Example: In a supermarket dataset:

- Rule: IF {Bread, Butter} THEN {Jam}
- Support = 0.15, Confidence = 0.65, Lift = 1.2

This means that 15% of all transactions contain all three items, and 65% of transactions containing Bread and Butter also include Jam.

Example Table: Frequent Itemset with Support and Confidence

Itemset	Support	Confidence
{Milk, Bread}	0.25	0.65
{Diaper, Beer}	0.10	0.70
{Eggs, Milk}	0.18	0.60

2. Model-Based Pattern Mining

Model-based pattern mining extracts interpretable patterns from **predictive machine learning**

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

models rather than directly from raw data. These approaches aim to leverage the predictive power of models while maintaining some level of human interpretability. By analyzing the internal structure of trained models, patterns and rules can be derived that **explain how inputs relate to outputs**.

Some widely used model-based pattern mining techniques include:

1. Decision Tree Induction

Decision trees are **inherently interpretable models** and widely used for pattern extraction. Each path from the root to a leaf node represents a **pattern or rule** that maps input features to a predicted output.

Example: Consider a healthcare dataset predicting diabetes risk:

- Root Node: Age > 50?
- Yes → BMI > 30?
- Yes → High Risk
- No → Medium Risk
- No → Family History = Yes?
- Yes → Medium Risk
- No → Low Risk

Each path in the tree can be represented as an **IF-THEN rule**, making it easy for practitioners to understand the model's decision logic.

Advantages:

- Directly interpretable without additional tools.
- Easy to visualize and explain to non-technical stakeholders.

Limitations:

- Can overfit on noisy data.
- Complex trees with many nodes reduce interpretability.

2. Rule Extraction from Neural Networks

Neural networks are powerful but **black-box models**, making their decisions hard to understand. Rule extraction techniques, such as **DeepRED**, **TREPAN**, and **RxN**, aim to convert the learned representations in neural networks into **symbolic IF-THEN rules**.

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Example Workflow:

- Train a neural network on a dataset (e.g., predicting loan defaults).
- Use DeepRED to extract rules:

Rule No	IF Condition	THEN Prediction
1	Age > 45 AND Income < 30k	High Risk
2	Age ≤ 45 AND Credit Score > 700	Low Risk
3	Age ≤ 45 AND Credit Score ≤ 700 AND Loan > 50k	Medium Risk

Advantages:

- Enables interpretation of otherwise black-box neural models.
- Useful in domains requiring **regulatory compliance** (e.g., finance, healthcare).

Limitations:

- Rule extraction may be computationally expensive for deep networks.
- Extracted rules may be approximate and not perfectly match the model.

3. Association Rule Discovery using Bayesian Networks

Bayesian networks (BNs) model **probabilistic relationships** among variables in a dataset. They can be used for pattern mining by interpreting **conditional dependencies** as patterns or rules.

Example: In a medical dataset:

- Rule: P(Diabetes | Obesity, Hypertension) = 0.78
- Interpretation: Patients with both obesity and hypertension have a 78% probability of developing diabetes.

Advantages:

- Incorporates uncertainty in patterns.
- Provides both **predictive power and interpretability**.
- Can be used for **causal analysis** if relationships are validated.

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Limitations:

- Learning large Bayesian networks can be computationally intensive.
- Conditional independence assumptions may not always hold.

3. Hybrid Approaches

Hybrid approaches aim to combine the strengths of **rule-based and model-based methods** to achieve both **high predictive performance and interpretability**. The idea is to use powerful predictive models for **pattern discovery** and then transform their results into **human-readable rules**.

1. Workflow of Hybrid Pattern Mining

- **Train a Predictive Model:** Use machine learning models such as **random forests, gradient boosting, or neural networks** to identify candidate patterns.
- **Pattern Extraction:** Apply symbolic or rule extraction techniques to convert model predictions into interpretable patterns.
- **Validation:** Filter and refine extracted rules using **support, confidence, and domain knowledge** to ensure meaningfulness.
- **Presentation:** Present the patterns in a human-readable format (tables, decision trees, or graphs).

Example: Using a random forest to predict customer churn:

- Train random forest on customer data.
- Extract rules using tree path aggregation:

Rule No	IF Condition	THEN Prediction
1	Contract = Month-to-Month AND Service Calls > 3	Churn
2	Contract = One-Year AND Payment Method = AutoPay	No Churn
3	Tenure < 6 Months AND Internet Service = Fiber	Churn

Here, the random forest provides **robust prediction**, and extracted rules provide **explanations** for business analysts.

2. Advantages of Hybrid Approaches

- Combines **high accuracy** of complex models with **interpretability** of rules.

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

- Supports **human-in-the-loop analysis**, allowing domain experts to validate patterns.
- Flexible: can be applied to various data types (tabular, sequential, or graph data).

3. Limitations of Hybrid Approaches

- Rule extraction may not capture all model complexities.
- Computationally more expensive than purely rule-based methods.
- Requires careful **pattern selection** to avoid overwhelming the user with too many rules.

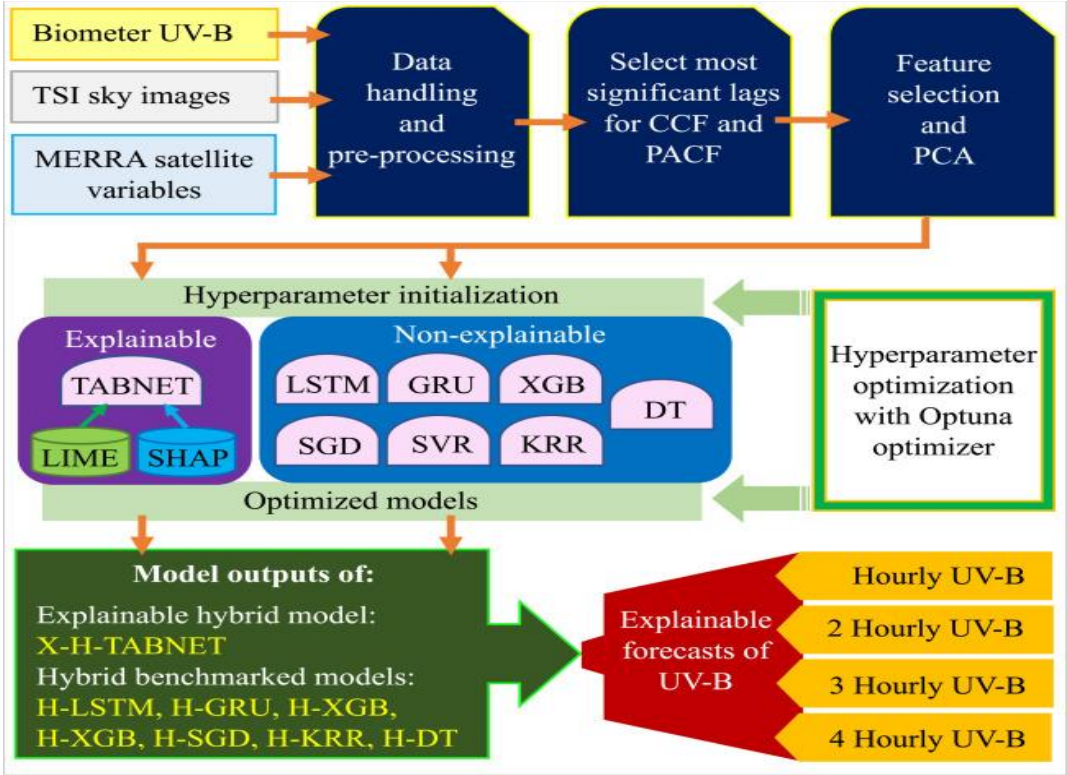


Figure 1: Hybrid Explainable Pattern Mining Pipeline

INTERPRETABLE MODEL TECHNIQUES

Interpretable models are designed to make machine learning predictions understandable to humans. They balance **predictive performance** with **transparency**, which is crucial in domains like healthcare, finance, and law, where decisions need justification.

1. Linear and Generalized Linear Models (GLMs)

Linear models are among the simplest and most interpretable approaches:

Key Features:

- The prediction is a weighted sum of input features.

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

- Coefficients indicate the **direction** (positive/negative) and **magnitude** of a feature's influence.
- Easy to visualize with plots like coefficient bar charts or partial dependence plots.

Example Applications:

- **Logistic regression**: Predicting whether a patient has a disease based on age, cholesterol, and blood pressure.
- **Linear regression**: Predicting house prices using square footage, number of bedrooms, and location.

Advantages:

- Highly interpretable and explainable.
- Efficient to train on large datasets.
- Well-suited for understanding feature effects.

Limitations:

- Assumes **linear relationships** between features and output.
- Struggles with capturing complex interactions unless explicitly modeled.

2. Decision Trees and Rule-Based Models

Decision trees and rule-based systems are structured approaches that provide **human-readable decision paths**.

Key Features:

- **Decision Trees:** Hierarchical structure where each node splits the dataset based on feature values.
- **Rule-Based Models:** Represent decisions as IF-THEN statements.

Example Applications:

- **Decision trees:** Credit scoring – approving or denying loans based on income, debt ratio, and payment history.
- **Rule-based models:** Medical diagnosis rules, e.g., IF fever > 101 AND cough = yes THEN flu risk = high.

Advantages:

- Intuitive visualization of predictions.
- Can explain individual decisions clearly.
- Works well with categorical and continuous data.

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Limitations:

- Prone to **overfitting**, especially on small datasets.
- May produce **unstable trees** with small changes in data.
- Limited ability to model complex non-linear relationships without ensemble methods.

3. Additive Models

Generalized Additive Models (GAMs) extend linear models by allowing **non-linear transformations** of features while preserving interpretability.

Key Features:

- Model output is a **sum of non-linear functions** of features: y=β0+f1(x1)+f2(x2)+...+fn(xn)y = \beta_0 + f_1(x_1) + f_2(x_2) + ... + f_n(x_n)y=β0+f1(x1)+f2(x2)+...+fn(xn).
- Each feature's contribution is **independent** and can be visualized.

Example Applications:

- Predicting hospital stay length using non-linear relationships with age, lab values, and vitals.
- Modeling risk scores in insurance with smooth effects of continuous features.

Advantages:

- Captures non-linear effects without sacrificing transparency.
- Feature contributions are easy to interpret and plot.

Limitations:

- Assumes **additivity**; cannot model complex feature interactions directly.
- More computationally intensive than simple linear models.

4. Post-hoc Explainability Techniques

For complex "black-box" models like **deep neural networks** or **gradient-boosted trees**, post-hoc methods help **explain predictions** after training.

Popular Techniques:

A. **LIME (Local Interpretable Model-agnostic Explanations):**

- Generates a simple surrogate model around a single prediction.
- Explains why the model made a specific prediction locally.

B. **SHAP (Shapley Additive Explanations):**

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

- Uses game theory to assign each feature a **contribution value** for a prediction.
- Can provide both local (single prediction) and global (overall feature importance) insights.

C. **Counterfactual Explanations:**

- Suggests minimal changes to input features to achieve a different prediction.
- Example: "If income increased by $5,000, loan approval would change from reject to approve."

Advantages:

- Makes highly complex models partially interpretable.
- Provides actionable insights for decision-making.

Limitations:

- Explanations may be approximate and not fully faithful to the original model.
- Can be computationally expensive for large datasets or complex models.

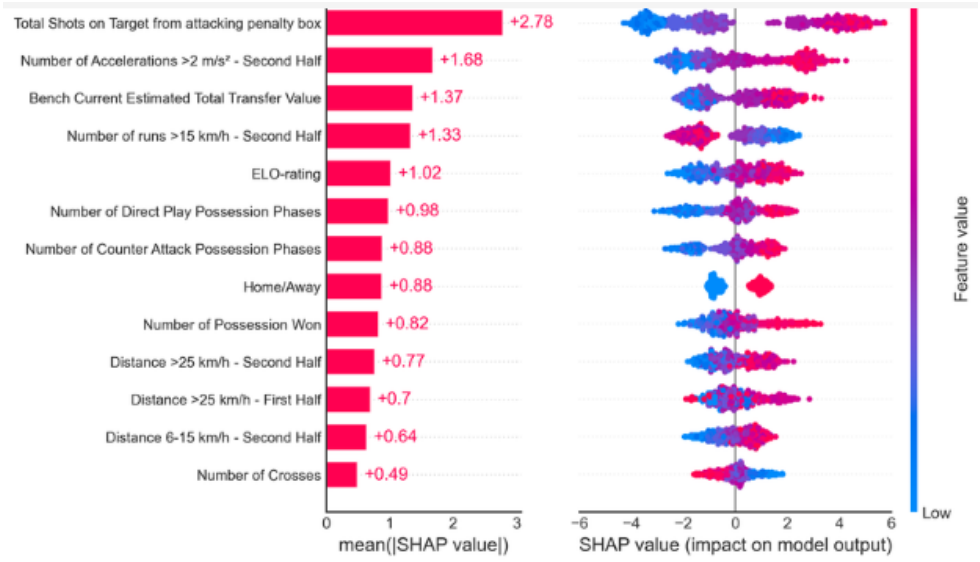


Figure 2: SHAP Explanation of Feature Importance

APPLICATIONS

1. Healthcare

- Predicting disease risks using interpretable models like decision trees and GAMs.
- Extracting patterns from electronic health records to discover treatment pathways.

2. Finance

- Fraud detection using explainable rules for suspicious transactions.
- Credit scoring models interpreted using SHAP to satisfy regulatory requirements.

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

3. Retail and Marketing

- Customer segmentation and recommendation systems using pattern mining.
- Example: Association rules identify items frequently purchased together.

4. Cybersecurity

- Detecting anomalies and attacks using interpretable models to explain alerts.
- Graph pattern mining to detect network intrusion patterns.

CHALLENGES AND LIMITATIONS

Despite progress, explainable pattern mining and interpretable models face challenges:

- **Trade-off between Accuracy and Interpretability:** Highly interpretable models may have lower predictive performance compared to deep learning.
- **Scalability:** Rule-based methods may struggle with large or high-dimensional datasets.
- **Ambiguity in Human Understanding:** Even "interpretable" patterns may be misinterpreted.
- **Temporal and Dynamic Data:** Maintaining explainability in dynamic environments (e.g., streaming data) is difficult.
- **Evaluation of Explainability:** No standardized metrics exist for measuring human-understandable explanations.

FUTURE DIRECTIONS

- **Integration with Knowledge Graphs:** Leveraging structured knowledge to enhance pattern interpretability.
- **Interactive Explainability Tools:** Interfaces that allow users to query models for explanations.
- **Explainable Deep Learning:** Developing deep learning architectures that are inherently interpretable.
- **Cross-Domain Pattern Mining:** Techniques that extract transferable patterns across domains.
- **Human-in-the-Loop Approaches:** Combining expert feedback with automated pattern mining to improve explainability.

CONCLUSION

Explainable pattern mining and interpretable models play a crucial role in bridging the gap

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved

between data-driven predictions and human understanding. Rule-based methods, model-based approaches, and hybrid techniques provide ways to discover meaningful and interpretable patterns. Interpretable models, including linear models, decision trees, and GAMs, offer transparent decision-making frameworks. While challenges such as scalability, evaluation, and trade-offs between accuracy and explainability remain, ongoing research is promising. Future work in interactive, knowledge-driven, and deep learning-based explainable approaches will enhance trust and adoption of machine learning in sensitive domains such as healthcare and finance.

REFERENCES

1. Han, J., Kamber, M., & Pei, J. (2021). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
2. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *KDD*.
3. Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *NIPS*.
4. Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD*.
5. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37–54.
6. Molnar, C. (2022). *Interpretable Machine Learning*. Lulu.com.
7. Duivesteijn, W., et al. (2016). Contrast Set Mining: Discovering Meaningful Differences between Groups. *Data Mining and Knowledge Discovery*.
8. Zeng, X., Chen, C., & Liu, X. (2020). Explainable AI: A Review of Methods and Applications. *IEEE Access*.
9. Wang, F., et al. (2019). Interpretable Machine Learning for Healthcare: Concepts, Methods, and Applications. *Annual Review of Biomedical Data Science*.
10. Guidotti, R., et al. (2018). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, 51(5).

Page 90-104 © MANTECH PUBLICATIONS 2016. All Rights Reserved