Metadata Causal Inference of Concept Drifts in Probabilistic- Relational Machine Learning: Predictive Analytics Using Graph Theory Methods in Cyber-Defense

**Fuzzy ML Concepts**

Karriem “A.J.” Perry**1\***

Barbara S. Keary**2**

Philip Kulp**1**

**1**Capitol Technology University

Doctoral Programs

11301 Springfield Road

Laurel, MD 20708 USA

{**\***kaperry[}@captechu.edu](mailto:%7d@captechu.edu);

**\***Corresponding Author

Emory University

Atlanta, GA USA

{b.keary}@outlook.com

**Disclosures:** The research, findings, and opinions expressed in this manuscript are solely those of the authors.

***Abstract*-**This research describes a novel approach to implementing cyber-defense protocols using supervised statistical-relational machine learning and lifted-inference in neural network architectures to detect concept drift anomalies. Data concept drift, or alternatively, concept shift, is a common yet difficult abnormality to detect in most data. Particularly metadata from adversarial cyber-attacks data originating from non-stationary environments, as detailed in the introduction. The Material and Methods and the Related Works sections describe the manuscript’s foundation in statistical-relational machine learning, otherwise known as relational machine learning, and its intrinsic suitability for identifying variables that may contain attributes, objects, entities and the like which contribute to concept drift irregularities during Exploratory Data Analysis (EDA) or Graphical Data Analysis (GDA) in detection of Structured Query Language Injection Attacks (SQLIA) probability. Moreover, balancing the tradeoffs when developing, Directed Acyclic Graphs (DAGs), Bayesian Networks (Bnets) and other network illustrations, demand increased computational time complexity. Lifted inference and prior probabilities are then introduced by the reproducible Analytics Solutions Unified Method for Data Mining/Predictive Analytics (ASUM-DM) framework, where handcrafting is efficacious and useful. The use of ASUM-DM also increases probable concept drift anomaly detection; predictive cyber defensive applications in chaotic systems and projected prior dependencies from lifted inference for datatypes originating from non-stationary environments. In later portions of this research, we describe how this approach allows for more efficient neural network design while enabling more consistent predictive analytics methods, enhancing proactive cyber-defense capabilities using modified activation functions amenable to a range of neural network architectures as a reflection of the Bayesian Information Criterion’s (BIC) output. This preemptive observation of the BIC value during epochs of neural networks identifies potential computational intractability and significantly reduces energy expenditure. We finalize hypothesis testing, neural network parameterization, and fine-tuning and develop a comprehensive evaluation and description of the findings, conclusion, and recommended areas of future research.

***Keywords:*** relational machine learning, concept drift, cyber-defense, SQL injection attack detection, nonparametric representations

# **Introduction**

This research combines the fields of metadata causal inference and statistical-relational machine learning and is focused on understanding the impact of concept drifts, which are alterations in the underlying data distribution, on the performance of statistical-relational predictive models for cyber-defensive applications (Kolter & Maloof, 2006). Concept drifts can significantly affect the accuracy of machine learning models and thus their study is an important area of research (Gama & Kosina, 2014). In the domain of cyber-defense, predictive analytics using graph theory methods is widely employed to identify potential security threats and vulnerabilities (Zhang & Zhang, 2011).

Statistical Relational Learning (SRL), alternatively referred to as Relational Machine Learning (RML), places special emphasis on metadata characteristics of variables in a given class of dependent and independent variables [1]. This study is focused on *structural* as opposed to *administrative* metadata; however, characteristic of the data under experimentation can function as *descriptive* metadata [1, 2]. These metadata objects, entities, elements, etc., typically compose the feature details of variables within the aggregate dataset as metadata [3]. Moreover, RML probabilistic reasoning that leverages a posteriori logic in learning approaches benefit a broader range of Machine Learning (ML) learning approaches and objectives [1, 2, 5]. Graphical probabilistic advances that employ these ML approaches include natural language processing; classification; computer vision, and many other current and emerging ML objectives (Theodoridis, 2015).

Graphical representations of latent variable entities, according to researcher (Nickle, 2015) can result in semantically structured information that is interpretable by computers illustrating relations among what is initially observed as otherwise unrelated variables properties (Chen et al., 2015; Sigtermans, 2021). It is important to make a distinctions between *dependence* and *causality,* and origins of these phenomena regardless of the supposition supporting, for example, derivatives of statistical dependencies (Geffner et al., 2022).

Graph theory functions (Dawood, 2014) applicable to the cybersecurity domain exhibit myriad utility specifically in illustrating potential joint distributions of Probability Density Functions (PDF); (Theodoridis, 2015). As the probability of adversarial cyber-attacks increase, cyber-actor behavior is progressively sophisticated and destructive and requires a more robust predictive and prescriptive analytics and analyses (McCallam et al., 2021). We propose to enhance the current utility of the computationally intensive methods, for instance Bayesian Attack Graphic (BAG) methodology (Matthews et al., 2020) by introducing a theorical Bayesian optimization method based on empirical observation discovered within of latent variables, or *metadata;* thereby increasing evaluation autonomy and reducing probable computational intractability resulting in communication, query, or communication estimations of problems ranging from *P* to *NP-*hardness (Angelov, 2019; Rubinstein, 2019).

# **Materials and Methods**

At the foundation of this research is Bayes’ rule, alternatively known as the *knowledge equation* (Lipovetsky, 2021), which often incorporates Bayesian optimization methods (Theodoridis, 2015) for instance, maximum likelihood functions (Chen et al., 2008; Gao et al., 2010), have shown increased utility by employing variations of the Bayesian Information Criterion (BIC) beyond the traditional means of model selection into larger aspects, as unit of measurement, for greater estimation of model approximations from graphical modeling (Chen et al., 2020; Schwarz, 1978; Theodoridis, 2015). Other popular and emerging statistical techniques involving Ordinary Least Squares (OLS), Beta regression, hybrid approaches, and stricter Bayesian methods leveraging, for instance, Markov Chain Monte Carlo (MCMC) testing procedures (Schulze et al., 2021).

However, these methods are irrevocability beholden to some level of structural vulnerability until the completeness of various observable states of the metadata are understood and structural assumptions of these data are minified (Geffner et al., 2022). To an extent, these structural assumptions are subordinate to the bounds of computational, query, and communication complexities and should be considered preemptively and concurrently as potential triggers to concept drift anomalies (Geffner et al., 2022; Rubinstein, 2019). For example, an immediate trigger indicating anomalous activity could be as simple as the use of a probabilistic tool identifying potential Virtual Private Network (VPN) inbound activity (Grechishnikov et. al, 2019). These lapses in accurate identification of relational elements within these data structures and potential anomalous activity or events has a higher probability of occurrence (Geffner et al., 2022; Koller, 2007).

## **Related Works**

Current literature addressing concept shift anomalies is largely concerned with streaming and time-series data types (Folino et al., 2019; Hashmani et al., 2019; Zhang et al., 2021; You et al., 2021). These data are typically of stationary origin and philosophically deterministic as opposed to probabilistic (Golden, 2020; Sugiyama et al., 2012). Recent prose are increasingly addressing graph theory approaches for a predictive analytics optic (Kulp et al., 2020). However, the degree of expansion on the subjects does not address forms of dataset shift (Quionero -Candela et al., 2008) or concept drift. AI and cyber security experts must seek adaptive and robust measures to counter an increasingly sophisticated cyber threat attack (Rubinstein, 2019), particularly with the introduction of a growing number of Internet of Things (IoT) devices and the unpredictability of how many of these devices respond to adversarial activity (Geng, 2017). Ostensibly, the *equation of knowledge* subjectively illustrates a unifying position to the larger philosophy of science, proving particular utility in the physical science leveraging a clear Bayesian perspective; postulations which allude to alternative theories (Lipovetsky, 2021). The researcher Lipovetsky (2021) demonstrates this a priori position in the following equation:

(1.0)

where Alter denotes all the alternative theories. From the prior , with the likelihood and the partition function which equals the expression in the denominator, the posterior probability of the Theory supported by data yields. (pg.140)

As indicated in Bayesian logic, updating one’s belief system with all possible alternative is an impossibility (Theodoridis, 2015). Further, the researcher Lipovetsky (2021), presents these alternatives illustrated in the following:

, (1.1)

where *T* and *D* denote the *[Theory]* and *[Data]*, respectively. When the new data News have been obtained independently from the past data *D*, the Bayesian inference reduces to

. (pg. 140); (1.2)

This postulation suggests that both priors and alternatives are probabilities applied from a Bayesian perspective (Lipovetsky, 2021). If modeling performance outputs the lowest BIC, the model’s architecture possess the most potential for accurate classification [7; Sugiyama et al., 2012). However, this measurement is calculated during training epochs independent of a formalized confusion matrix; F1 score; Receiver Operating Characteristic (ROC), and so on (Du & Swamy, 2019).

Now, let us discuss the types of probabilistic graphical models of interests. These models are:

1. conditional Bayesian networks
2. temporal Bayesian networks
3. multi-entity Bayesian networks

Instances where model selection is

## **Research Problem Statement**

Many data analysis techniques used in cyber defense do not incorporate the Analytics Solutions Unified Method for Data Mining/Predictive Analytics (ASUM-DM) methodology or alternate approaches used in process control (Angée et al., 2018; Erskin, 2010). This absence in using a comprehensive analytics methodology across the cyber defense domain leads to inconsistencies in developing ubiquitous cyber security strategies and technique reproducibility in a larger data science strategy (Angée et al., 2019). Considering the Pareto Principal, alternatively the 80/20 rule; this absence of a comprehensive and distinguishable set of methods compounds the preexisting complications generally accompanying Exploratory Data Analytics (EDA) procedures (Angée et al., 2018; Harvey et al., 2018).

## **Concept Drift Defined**

Several forms of dataset shift anomalies may contribute to the presence of an identifiable concept drift event, chiefly of data originating from non-stationary environments, for instance, dynamic data streams (Kokilam et al., 2020; Liu et al., 2021; Quionero-Candela et al., 2008). This type of dataset shift can be defined as (Barbero, 2020) where the probability ( is consequently affected by no longer producing reliable model approximations in classification tasks. Other variations of shift anomalies impairing posterior probabilities and input distribution identified as *real drifts* and *virtual drifts*, respectively (Zhang et al., 2021). Considering when the element proves problematic and by proxy significantly degrading the model’s learning rate (Brownlee, 2019), accurately estimating the characteristics of these elements’ objects and/or entities that comprise the metadata is paramount in determining the true structure of these variables under training and testing scenarios and crucial to estimate accuracy (Koller et al., 2007; Sugiyama et al., 2012; Theodoridis, 2019).

Examining the root cause(s) of probable shift events, whether induced by an unexpected function of time; misinterpretation of data structures or arrays whereas seminal characteristics typically exist within these often-complex values signaling the potentiality of drift anomalies occurring (Bellot, 2016; Quionero-Candela et al., 2008). As with random variables, these events are measured using probabilities (Bellot, 2016) and serve as estimations of the occurrence of anomalous events which otherwise would exist unbeknownst during model development. This notion sufficiently links probability theory and possibility theory as conditions of joint or conditional probability rooted in first-order logical belief systems and identification of the resultant conditions of uncertainty (Bellot, 2016; Carlsson et al, 2011). Furthermore, these juxtapositions illustrate the transition from fuzzy to calculable interpretations in Bayesian optimization (Bellot, 2016; Theodoridis, 2019). These quantifiable clarifications are further classified in the scope of *probability calculus* (Bellot, 2016) and assist in describing the statistical architecture of methods in optimization for dataset shifts (Quionero-Candela et al., 2008; Theodoridis, 2019).

#### **Concept Drift Predictability**

From a cyber defense perspective, changes in characteristics, patterns, types, etc., of unexpected change constitute *concept drift* (Folino et al., 2019)*.* Central to accurately identifying potential concept drift and subsequently apply remediation measures, it is necessary to understand and define the scope of ML approached projected in model development. In the context of graphical networks which enhance illustrations of relationships (Koller, 2007) in Bayesian optimization methods (Theodoridis, 2015), we will focus primarily on semi-supervised learning applications which directly reflect outputs of semi-supervised ML estimates and approximations (Quionero-Candela et al., 2008). Presuming the training datasets are Independently and Identically Distributed (i.i.d.) and do not originate from an environment of non-stationarity (Sugiyama et al., 2012); therefore, a significant probability exists that these datasets will contribute concurrently contribute to developing the optimization techniques needed to promote model convergence (Aggarwal, 2018; Quionero-Candela et al., 2008). The exception to i.i.d. persists in some Bayesian Model Comparison (BMC) methods especially when used in change-point detection (Bach et al., 2010).

Following identification of the composition of these metadata and discovering the simple structure of those data becomes essential to the learning conditions of the model (Lane, 1998). We have identified several viable Bayesian optimization techniques, using EDA and GDA methods covered in later sections (Bellot, 2016), which show promising results in effectively countering concept drift irregularities in non-stationary data (Adams, 2014). However, these concept drift abnormalities are not exclusive to class imbalance in streaming data which typically are responsible , for instance, Type I and/or Type II errors in confusion matrices (Wang, 2018) or from data of non-stationary origin (Lu et al., 2019; Quionero-Candela et al., 2008).

The influences and conditions of metadata relations affecting variables may vary on a narrow to broad spectrum of probabilities. These probabilities are attributed to *latent variable spaces* which can describe attributes, objects, or elements within variables prior to EDA, GDA, or statistical testing (Bellot, 2016; Xu, 2021). Particular to the scope of statistical inference while consider random vector X; typically, a given random variable contains the unknown probability density functions . Our analysis assumes that one or more of the sampled population distributions will illustrate some characteristic of stochastic dominance, whether *first, second, third* or *higher-order* canonical or otherwise (Birnbaum, 2005; Chen, 2020; Müller, 2017; Niu, 2017). This belief thereby requires preemptive test of variables metadata using the Wilcoxon Mann-Whitney *U-*statistic test to accept or reject the null hypothesis as a process of determining the probability of concept drift or dataset shift activity (Mangiafico, 2016; 7]. This this *U-*critical or *U-*statistical outcome assists in defining vector importance, a crucial aspect in determining, for example, *eigenvector centrality* (Lee, 2015; Khwaja, 2021).

### **Lifted Probabilistic Utility**

Lifted probability, is being identified as more prevalent due to scientific data and the growing number of organizations and institutions in the public/private; academia and industry (Van den Broeck, 2011). Thorough understanding of the structural characteristics of those data under evaluation, and subsequent metadata analysis, is requisite to graphically representing relations among covariates (Koller, 2012). A critical aspect in determining conditions of the attributes, objects, entities, and other forms of metadata, is the development of schemas and mapping scenarios which accurately outline the structure within variable metadata (Van den Broeck et al., 2021). This is accomplished by approximating variable distributions and examining metadata for structures consistent between covariates [7]. In tandem with EDA, development of graphical models is performed developing nodes and edges of the most appropriate directed or undirected graphical model (Bellot, 2016; Khwaja, 2021).

Bayesian Networks (Bnet), a type of directed graph, retains a specific purpose and task. Considering both directed and undirected graphs as representations of set theory when building abstractions of complex models prior to defining, for instance, a *d-*dimensional feature vector (Golden, 2020). A classic example of a given vector space can be described as:

. (2.0)

This *standard basis* can be further simplified, when defining the *finite-dimensional* vector space where represents the set of all non-negative, real values, as over a given *field S* as ; (Itskov, 2009). We define the vector space and field to represent *time* and *stationarity* respectively in this research.

For context, it is important to conceptualize the true nature of a given random variable, which contains an unknown *first moment*, or, which is defined as devoid of the descriptive characteristics of its nomenclature (Liu, 2010; Van den Broeck et al., 2021). Further for the purposes of clarity and consistency, this expectation is defined as:

,. (2.1)

The assumption, in Fig 2.1, is the discriminator in this definition; since approaches in, for instance, Borel functional calculus, spectral measurement, or conditions describing *pure states* in conjunction to a vector derived from classical probability theory may seem somewhat arbitrary from casual observation; however, these characteristics takes on particular significance when conditions of uncertainty exist, in the context of this work, that are categorized as non-stationarity and we are only concerned with circumstances highlighted in Fig 2.1 (Liu, 2010; Sugiyama et al., 2012).

### ***Bayesian Classification + (Plus)***

Unlike other classification models, for instance support vector machines; a Bayesian which includes Gaussian possess self-contained Automatic Relevance Determinator (ARD); (Williams et al., 1998). Extending BIC utility into random variables high-dimensional coverage, we consider graphical models’ inference of Multi-classification (*M-*class) problems as opposed to the binary *Ising model* approach (Foygel et al., 2011; Perry, 2021). This local instance, as described in terms of Explainable Artificial Intelligence (XAI) from a *first-order* logic-based optic, provides foundational workings in the probabilistic sphere using a modified BIC as the unit of measurement in model performance evaluation (Wang et al., 2021). Consider the standard parametric BIC theory, in which typically a sample size of observations whereas the model possesses the log-likelihood function = . Therefore, the typical form of BIC for this particular model, , is:

BIC (3.0)

As indicated, represents the finite dimension of the parameter space, which poses a problem when considering the probable infinite-dimensionality represented in an *M-*class vector space and subsequent model’s parameter space signifying the Maximum Likelihood Estimator (MLE):

; (Wang et al., 2021). (3.1)

The non-stationary properties of *M-*class or binary problems in this research compel the substitution of the BIC classical theory (Wang, 2021) with a modified BIC approach equivalent to the following theorem:

**Theorem 1.** Letbe the estimate of the training or testing epoch of a given neural network cycle. In this instance, epochs are under evaluation for an *M-*class or binary problem. The data are from non-stationary origin. Meaning, the BIC could serve as an early indicator of model performance shown in the below proof (Sugiyama et al., 2012):

**Proof 1.** , where ; (Caldera. 2009; 7]. (4.0)

Therefore, to determine the extent of the relations among covariates and how these relations may influence model performance during the initial training epoch; and more importantly the impact on the composition of the data structure and establish if correlation or causation exists, or variations thereof (Van den Broeck et al., 2021). Therefore, we will employ lifted causal inference as the primary statistical method for hypothesis testing (Van den Broeck et al., 2021).

1. **Predictive Cyber Defense**

In many instances, cyber applications are targeted using Structure Query Language Injection Attacks (SQLIA); (Archana et al., 2021). Current ML cyber-defense methods often employ modeling parameters which evaluate information under the aegis of uncertainty, specifically probability and possibility theory, fuzzy logic and its subordinate domains in uncertainty theory*; credibility;* and *chance theory* (Liu, 2010; Archana et al., 2021). The *randomness* and *fuzziness* intrinsic to uncertainty this uncertainty is a derivation of the variable structures’ *implicit membership* and is often a consideration unbeknownst to many cyber-security professionals (Wang, 1994).

Determining the level of membership and subsequent relational components must be considered from a range of possibilities (Liu, 2010). Anticipating and correctly identifying these possible shift anomalies of concepts is impossible deduce with some level of minimal assumption(s); (Liu, 2010; Sugiyama, 2012).

1. **Research Questions**

This study is composed of two research questions:

1. *Research Question 1 (RQ1): How does the incidence of concept drifts impact the predictive accuracy of statistical-relational machine learning models in the domain of cyber-defense?*
2. *Research Question 2 (RQ2): Can graph theory methods be utilized to detect and mitigate the impact of concept drifts on the predictive accuracy of statistical-relational machine learning models in cyber-defense?*

Hypothesis 1: The predictive accuracy of statistical-relational machine learning models will experience a decrease as the frequency and magnitude of concept drifts increase.

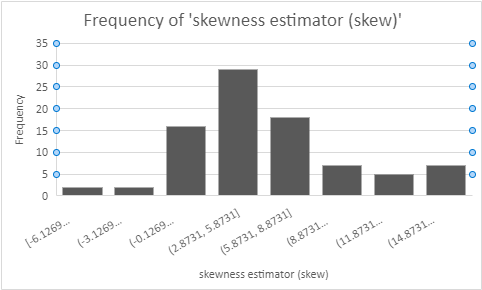
Hypothesis 2: The utilization of graph theory methods will result in an improvement in the predictive accuracy of statistical-relational machine learning models in the presence of concept drifts.

1. **Data analysis**

Using the NATICUSdroid (Android Permissions)dataset (Mathur, 2021), with binary values zero and one, containing 29,333 instances and 86 variables and one target variable. The disposition of these datapoints is static and does not appear to change the characteristics of variables from casual observations of the variables. The number of attributes, equaling 115, are evaluated during training epoch(s) using modified versions of two separate activation functions. These functions include the Sigmoid Linear Unit (SiLU), which showed utility in previous research [7]. In addition, due to the size of the dataset, the TanH activation function is also employed to compare multiple BIC values in parallel to activation functions under the same neural network architecture. This empirical approach is assuming that some degree of shift/drift anomaly persist within those metadata elements, objects, or entities and the numerical values of the BIC outputs serves as an indicator of a given neural networks probable performance (Sugiyama et al., 2012).

The dataset contains 86 covariates and one target variable, *results.*  We will use d'Agostino (1971) tests of skewness which assists determining the skewness and kurtoic properties of the data (d’Agostino, 1971). Figure 1 illustrates the need to show potential relationships among variables as we will discover, for example, how one covariate may influence the exponential growth or potential linearity of distribution over a sequenced event

**Figure 2. Estimated Skewness**

Chart

Description automatically generated

(a) Estimate of frequency skewness (b) Frequency transformation

*Note**:* These approximations show the frequency relative to observed skewness of the data. However, this postulation does not completely illustrate the data characteristics and only alludes to a semblance of i.i.d.

**Figure 3. Frequency after Transformation**

*Note**:* The frequency distribution transformed to a more positive skew which better reflects the characteristics and potential performance of these data.

* 1. ***Exploratory Data Analysis (EDA)***

In statistical analysis, principal component analysis (PCA) and factor analysis (FA) are two commonly used methods for data reduction and dimensionality reduction [5]. While PCA aims to extract a set of linearly uncorrelated components that account for the maximum amount of variance in the original data, FA aims to identify a set of latent factors that underlie the observed covariates [6].

The simple structure of the covariates may contain latent data; therefore, we use Principal Factor Analysis (FCA), as opposed to Principal Component Analysis (PCA) to determine the composition of the variables and if these latent structures [6]. In addition, PCA alone does not differential among types of variance account for all forms of variance [5]. This initial depiction is displayed in Figure 4. In addition, PCA alone does not differentiate among the types of variances and may not account for all forms of variance present in the data. According [5], PCA tends to extract the most strongly correlated variables and ignores the weaker correlations. This limitation can be overcome by using FA, which can capture more complex relationships among the variables and identify underlying factors that explain the observed covariance [5, 6].

**Figure 4. Principal Factor Analysis Screenplot**

Chart, bar chart

Description automatically generated

(c) Estimate of frequency skewness

*Note:*  There are ten listed components from the manifested variables. However, these components may be transformed to no more (>) or less than (<) three factors.

In identifying the composition of the dataset, it is common practice to split the data into training and testing sets to evaluate the performance of the model on new, unseen data. This approach can help to identify overfitting, which occurs when the model learns the training data too well and fails to generalize to new data. In this study, the dataset was split using a 70/30% split into training and testing samples, respectively. This split ensures that there is sufficient data for the model to learn from while also providing a large enough sample for testing.

Once the training and testing datasets were established, the statistical-relational machine learning models were trained and tested using the SRL toolkit, Alchemy (Kok & Domingos, 2005). Alchemy is a widely used toolkit for statistical relational learning that supports a range of probabilistic modeling languages, including Markov logic networks, which are particularly well-suited for modeling relational data.

The trained models were then evaluated using a range of metrics, including precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve. These metrics provide a comprehensive evaluation of the performance of the models and can help to identify areas for improvement. The results of the evaluation were used to fine-tune the models and optimize their performance for the given task.

Kok, S., & Domingos, P. (2005). Learning the structure of Markov logic networks. Proceedings of the 22nd international conference on Machine learning, 441-448.

Nisbet, R., Elder, J., & Miner, G. (2009). Handbook of statistical analysis and data mining applications. Academic Press.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big Data, 1(1), 51-59.

**Figure 6. XGBoost Feature Importance (training)**A picture containing text

Description automatically generated

(d) Feature importance\_training data

*Note:* Feature importance clusters six and 7 show illustrate the most influence over the structure of the composition of the nodes, leaves, and forests of the training model.

**Table 1. Feature Matrix (training)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **Gain** | **Cover** | **Frequency** | **Importance** |
| android\_permission\_READ\_PHONE\_STATE | 65.93% | 19.85% | 3.09% | 65.93% |
| com\_google\_android\_c2dm\_permission\_RECEIVE | 19.27% | 11.74% | 3.09% | 19.27% |
| android\_permission\_KILL\_BACKGROUND\_PROCESSES | 2.03% | 5.30% | 1.85% | 2.03% |
| android\_permission\_ACCESS\_FINE\_LOCATION | 1.70% | 1.87% | 4.32% | 1.70% |
| android\_permission\_RECEIVE\_BOOT\_COMPLETED | 1.59% | 5.94% | 3.70% | 1.59% |
| android\_permission\_INTERNET | 1.41% | 8.02% | 4.94% | 1.41% |
| android\_permission\_SEND\_SMS | 1.21% | 7.90% | 6.17% | 1.21% |
| android\_permission\_READ\_EXTERNAL\_STORAGE | 1.21% | 0.99% | 6.79% | 1.21% |
| android\_permission\_ACCESS\_COARSE\_LOCATION | 0.82% | 0.62% | 6.17% | 0.82% |
| android\_permission\_SYSTEM\_ALERT\_WINDOW | 0.82% | 1.06% | 5.56% | 0.82% |
| com\_android\_launcher\_permission\_INSTALL\_SHORTCUT | 0.78% | 2.18% | 4.94% | 0.78% |
| android\_permission\_GET\_TASKS | 0.61% | 0.44% | 1.85% | 0.61% |
| android\_permission\_WAKE\_LOCK | 0.54% | 0.37% | 4.94% | 0.54% |
| android\_permission\_REQUEST\_INSTALL\_PACKAGES | 0.47% | 6.99% | 3.09% | 0.47% |
| android\_permission\_RECEIVE\_USER\_PRESENT | 0.34% | 5.48% | 2.47% | 0.34% |
| com\_android\_vending\_BILLING | 0.32% | 4.56% | 3.09% | 0.32% |
| android\_permission\_RECORD\_AUDIO | 0.19% | 0.51% | 1.23% | 0.19% |
| android\_permission\_MODIFY\_AUDIO\_SETTINGS | 0.12% | 0.42% | 2.47% | 0.12% |
| com\_samsung\_android\_providers\_context\_permission\_WRITE\_USE\_APP\_FEATURE\_SURVEY | 0.11% | 4.26% | 1.23% | 0.11% |
| com\_sonyericsson\_home\_permission\_BROADCAST\_BADGE | 0.09% | 2.62% | 1.23% | 0.09% |
| android\_permission\_VIBRATE | 0.06% | 0.10% | 1.85% | 0.06% |
| android\_permission\_android\_permission\_READ\_PHONE\_STATE | 0.06% | 2.23% | 0.62% | 0.06% |
| android\_permission\_READ\_LOGS | 0.04% | 0.05% | 2.47% | 0.04% |
| android\_permission\_GET\_ACCOUNTS | 0.04% | 0.04% | 2.47% | 0.04% |
| android\_permission\_WRITE\_EXTERNAL\_STORAGE | 0.03% | 0.90% | 4.32% | 0.03% |
| android\_permission\_CALL\_PHONE | 0.03% | 0.21% | 1.85% | 0.03% |
| android\_permission\_USE\_FINGERPRINT | 0.02% | 1.03% | 0.62% | 0.02% |
| android\_permission\_READ\_CONTACTS | 0.02% | 0.01% | 0.62% | 0.02% |
| android\_permission\_MANAGE\_ACCOUNTS | 0.02% | 0.16% | 1.23% | 0.02% |
| android\_permission\_MOUNT\_UNMOUNT\_FILESYSTEMS | 0.02% | 0.00% | 0.62% | 0.02% |
| android\_permission\_RESTART\_PACKAGES | 0.02% | 0.03% | 0.62% | 0.02% |
| android\_permission\_ACCESS\_WIFI\_STATE | 0.02% | 0.04% | 1.23% | 0.02% |
| android\_permission\_NFC | 0.02% | 1.38% | 0.62% | 0.02% |
| android\_permission\_BROADCAST\_STICKY | 0.01% | 0.05% | 1.85% | 0.01% |
| android\_permission\_DOWNLOAD\_WITHOUT\_NOTIFICATION | 0.01% | 0.63% | 1.85% | 0.01% |
| com\_android\_vending\_CHECK\_LICENSE | 0.01% | 0.41% | 1.23% | 0.01% |
| android\_permission\_CHANGE\_WIFI\_STATE | 0.00% | 0.01% | 0.62% | 0.00% |
| com\_android\_launcher\_permission\_UNINSTALL\_SHORTCUT | 0.00% | 0.01% | 0.62% | 0.00% |
| android\_permission\_ACCESS\_NETWORK\_STATE | 0.00% | 1.20% | 0.62% | 0.00% |
| com\_android\_launcher\_permission\_READ\_SETTINGS | 0.00% | 0.24% | 0.62% | 0.00% |
| android\_permission\_READ\_PROFILE | 0.00% | 0.01% | 0.62% | 0.00% |
| android\_permission\_SET\_WALLPAPER | 0.00% | 0.13% | 0.62% | 0.00% |

**Figure 7. BIC-based Clustering Plot**Chart, line chart

Description automatically generated

*The*

**Figure 8. Population Clustering Plot** Chart, scatter chart

Description automatically generated

**Figure 9. Population Clustering Plot**

Chart, scatter chart

Description automatically generated

* 1. **Black Box Interpretation**

One of the central challenges in the current trajectory of ML is model interpretation and explainability (Kamath et al., 2021). These processes, deemed *XAI*, can be classified as mutually supporting yet exclusive in some regard (Kamath et al., 2021; Wang et al., 2021). Meta-analysis of feature using *gain* and *cover* will provide a perspective in covariate/feature structure.

* + 1. **Theoretical Relational Structure**

The theoretical structure of path models describes possible relationships among the

1. **Experimentation Setup**

The research in this study is developed using the R© statistical programming language, open-data community syntax (Team, 2017). Using this approach will allow maximum availability to both theoretical and applied research communities and promote reproducibility for potential meta-analysis projects. Once the

1. **Hypothesis Statement and Testing**

Prior knowledge of events is applicable to the development of the following statement, which has application in both *probability* and *possibility theory.*

### **Learning Rate Refinement**

Initiated with modeling training epochs are episodic learning rates. These learning rates are subject to change given the composition of the neural network(s) in training phase (Du et al., 2019). Moreover, this parameter is specific to the modeling architecture, size of the training, testing, and validation datasets, number of features/parameters, etc.; (Du & Swamy, 2019).

## **Optimization Procedures**

Many of the optimization issues associate

d with identifying and mitigating concept drift include the need for personalization, timeliness, and contextual information to provide to these models. Optimization is also made difficult because of the concept drift that occurs over time (Hashmani et al., 2019). Hashmani et al, (2019) found that concept drift may make deep learning networks degraded based on inconsistent classification data. Researchers like (You, et al., 2021) developed a framework to model concept drift during interference and remove model aging concerns with the data. But there are still many concerns around optimization of data with the ongoing concern over concept drift. However, there are still many concerns around optimization of data with the ongoing concern over concept drift anomalies (Sugiyama, 2012). Considering the *M-*class problems associated with non-normalized vector space predictions, logits often transform into inputs to a given activation function (Theodoridis, 2015), all of which precedes convergence (Du & Swamy, 2019).

## **Findings**

## **16. Conclusion**

## **16.1. Recommended Future Research**

# **References**

[1] Koller, D., Friedman, N., Džeroski, S., Sutton, C., McCallum, A., Pfeffer, A., ... & Yih, W. T. (2007). *Introduction to statistical relational learning*. MIT press.

[2] Svore, K. M., Bennett, P. N., & Dumais, S. T. (2015). *U.S. Patent No. 9,020,936*. Washington, DC: U.S. Patent and Trademark Office.

[3] Verger, T., & Robertson, S. (2008). GATS Basics: Key Concepts.

[3] Harring, J.R., Kohli, N., Silverman, R.D. and Speece, D.L., 2012. A second-order conditionally linear mixed effects model with observed and latent variable covariates. *Structural equation modeling: a multidisciplinary journal,* 19(1), pp.118-136.

[4] Frazier, P. I. (2018). A Tutorial on Bayesian Optimization. *stat*, *1050*, 8.

[5] Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2019). Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge & Data Engineering*, *31*(12), 2346-2363.

[6] Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four

recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1), 7.

[7] Karer, T., Steiner, P. M., & Wiedermann, W. (2012). Principal Factor Analysis versus Principal Component Analysis in Dataset Reduction. Proceedings of the 4th International Conference on Machine Learning and Data Mining in Pattern Recognition, 504-515.

[8] Perry, K. A. J. (2021). *Á Posteriori Bias Identification and Remediation of Dataset Shifts in Deep Structured Relational Machine Learning*(Order No. 28965169). Available from ProQuest Dissertations & Theses Global. (2638300084). <https://www.proquest.com/dissertations-theses/á-posteriori-bias-identification-remediation/docview/2638300084/se-2?accountid=44888>

Archana Devi, R., Amritha, C., Sai Gokul, K., Ramanuja, N., & Yaswant, L. (2021). Prevention and

Detection of SQL Injection Using Query Tokenization. In *Advances in Distributed Computing and Machine Learning* (pp. 165-172). Springer, Singapore.

Adams, R. P. (2014). A tutorial on Bayesian optimization for machine learning. *Harvard University*.

Aggarwal, C. C. (2018). Neural networks and deep learning. *Springer*, *10*, 978-3.

Angée, S., Lozano-Argel, S. I., Montoya-Munera, E. N., Ospina-Arango, J. D., & Tabares-Betancur,

M. S. (2018, August). Towards an improved ASUM-DM process methodology for cross-disciplinary multi-organization big data & analytics projects. In *International Conference on Knowledge Management in Organizations* (pp. 613-624). Springer, Cham.

Angelov, P. P., & Gu, X. (2019). *Empirical approach to machine learning.* Cham: Springer.

Bach, S., & Maloof, M. (2010). A bayesian approach to concept drift. In *Advances in neural*

*information processing systems* (pp. 127-135).

Baldi, P., Sadowski, P., & Whiteson, D. (2014). Searching for exotic particles in high-energy physics

with deep learning. *Nature communications*, *5*(1), 1-9.

Bellot, D. (2016). *Learning probabilistic graphical models in R*. Packt Publishing Ltd.

Birnbaum, M. H. (2005). A comparison of five models that predict violations of first-order stochastic

dominance in risky decision making. *Journal of Risk and Uncertainty*, *31*(3), 263-287.

Brownlee, J. (2019). Understand the impact of learning rate on neural network performance.

Machine Learning Mastery.

Carlsson, C., & Fullér, R. (2011). Possibility for decision. In *Studies in Fuzziness and Soft*

*Computing* (Vol.270). Springer.

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large

model spaces. Biometrika, 95(3), 759-771.

Cham. Foygel, R., & Drton, M. (2011). Bayesian model choice and information criteria in sparse

generalized linear models. *arXiv preprint arXiv:1112.5635*.

Chan, R. H., Clark, E., Guo, X., & Wong, W. K. (2020). New development on the third-order

Stochastic dominance for risk-averse and risk-seeking investors with application in risk management. *Risk Management*, *22*(2), 108-132.

Chen, Y., Meng, L., & Tian, J. (2015, February). Exact Bayesian learning of ancestor relations in

Bayesian networks. In Artificial Intelligence and Statistics (pp. 174-182).

Du, K. L., & Swamy, M. N. S. (2019). *Neural Networks and Statistical Learning*. Springer Nature.

Erskine, J. R., Peterson, G. L., Mullins, B. E., & Grimaila, M. R. (2010, April). Developing

cyberspace data understanding: using CRISP-DM for host-based IDS feature mining. In *Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research* (pp. 1-4).

Folino, G., Pisani, F. S., & Pontieri, L. (2019, June). A cybersecurity framework for classifying non

stationary data streams exploiting genetic programming and ensemble learning.

In *International Conference on Numerical Computations: Theory and Algorithms* (pp. 269-277). Springer,

Frezat, H., Le Sommer, J., Fablet, R., Balarac, G., & Lguensat, R. (2021). A posteriori learning of

quasi-geostrophic turbulence parametrization: an experiment on integration steps. arXiv e-prints, arXiv-2111.

Fridman, L., Weber, S., Greenstadt, R., & Kam, M. (2016). Active authentication on mobile devices via

stylometry, application usage, web browsing, and GPS location. *IEEE Systems Journal*, *11*(2), 513-521.

Geffner, H., Dechter, R., & Halpern, J. Y. (Eds.). (2022). Probabilistic and Causal Inference: The

Works of Judea Pearl.

Golden, R. M. (2020). *Statistical machine learning: A unified framework*. Chapman and Hall/CRC.

Grechishnikov, E. V., Dobryshin, M. M., Kochedykov, S. S., & Novoselcev, V. I. (2019, April).

Algorithmic model of functioning of the system to detect and counter cyber attacks on virtual private network. In *Journal of Physics: Conference Series* (Vol. 1203, No. 1, p. 012064). IOP Publishing.

Hashmani, M. A., Jameel, S. M., Alhussain, H., Rehman, M. and Budiman, A. (2019). Accuracy

performance degradation in image classification models due to concept drift. International Journal of Advanced Computer Science and Applications, 10(5), 2019.

Itskov, M. (2009). Tensor Algebra and Tensor Analysis for Engineers: With Applications to Continuum Mechanics.

Kamath, U., & Liu, J. (2021). Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning (pp. 1-310). Springer.

Karer, Gorazd, and Igor Škrjanc. Predictive approaches to control of complex systems. Vol. 454. Springer, 2012.

Khwaja, E. T. (2021). *The Network Architecture of Rural Development Interventions: Exploring the Relational Dynamics of Aid-Impact in the Fragile and Conflict-Affected States of Pakistan and Afghanistan* (Doctoral dissertation, George Mason University).

Kokilam, K. V., & Latha, D. P. P. (2020, December). Ensemble Method to classify multi class with concept drift. In *Journal of Physics: Conference Series* (Vol. 1706, No. 1, p. 012151). IOP Publishing.

Kulp, P. H., & Robinson, N. E. (2020, November). Graphing Website Relationships for Risk Prediction: Identifying Derived Threats to Users Based on Known Indicators. In *Proceedings of the Future Technologies Conference* (pp. 538-549). Springer, Cham.

Lane, T., & Brodley, C. E. (1998, August). Approaches to Online Learning and Concept Drift for User Identification in Computer Security. In *KDD* (pp. 259-263).

Lipovetsky, S. (2021). The Equation of Knowledge: From Bayes’ Rule to a Unified Philosophy of Science.

Liu, A., Lu, J., & Zhang, G. (2020). Diverse instance-weighting ensemble based on region drift disagreement for concept drift adaptation. *IEEE transactions on neural networks and learning systems, 32*(1), 293-307.

Liu, B. (2010). Uncertainty theory. In *Uncertainty theory* (pp. 1-79). Springer, Berlin, Heidelberg.

Liu, Y. H., & Ha, M. (2010). Expected value of function of uncertain variables. *Journal of uncertain Systems*, *4*(3), 181-186.

Mangiafico, S. S. (2016). Summary and analysis of extension program evaluation in R. *Rutgers Cooperative Extension: New Brunswick, NJ, USA*, *125*, 16-22.

Mathur, A., Podila, L. M., Kulkarni, K., Niyaz, Q., & Javaid, A. Y. (2021). NATICUSdroid: A malware detection framework for Android using native and custom permissions. *Journal of Information Security and Applications*, *58*, 102696.

Matthews, I., Mace, J., Soudjani, S., & van Moorsel, A. (2020, December). Cyclic Bayesian attack graphs: A systematic computational approach. In *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (pp. 129-136). IEEE.

Müller, A., Scarsini, M., Tsetlin, I., & Winkler, R. L. (2017). Between first-and second-order stochastic dominance. *Management Science*, *63*(9), 2933-2947.

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2015). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, *104*(1), 11-33.

Niu, C., Wong, W. K., & Xu, Q. (2017). Kappa ratios and (higher-order) stochastic dominance. *Risk Management*, *19*(3), 245-253.

Quinonero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (Eds.). (2008). *Dataset shift in machine learning*. Mit Press.

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., ... & Edelman, A. (2020). Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*.

Rossi, R. A. (2015). *Improving Relational Machine Learning by Modeling Temporal Dependencies* (Doctoral dissertation, Purdue University).

Rubinstein, A. (2019). *Hardness of Approximation Between P and NP*. Morgan & Claypool.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Sigtermans, D. (2021). Determining Causal Skeletons with Information Theory. Entropy, 23(1), 38.

Soetaert, K., Cash, J., & Mazzia, F. (2012). Solving ordinary differential equations in R. In *Solving Differential Equations in R* (pp. 41-80). Springer, Berlin, Heidelberg.

Sugiyama, Masashi, and Motoaki Kawanabe. *Machine learning in non-stationary environments: Introduction to covariate shift adaptation*. MIT press, 2012.

Team, R. C. (2017). R: a language and environment for statistical computing. R Foundation for

Statistical Computing, Vienna. https://www.R-project.org.

Van den Broeck, G., Taghipour, N., Meert, W., Davis, J., & De Raedt, L. (2011). Lifted probabilistic inference by first-order knowledge compilation. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence* (pp. 2178-2185). AAAI Press/International Joint Conferences on Artificial Intelligence; Menlo Park, California.

Van den Broeck, G., Kersting, K., Natarajan, S., & Poole, D. (Eds.). (2021). *An Introduction to Lifted Probabilistic Inference*. MIT Press.

Wang, E., Khosravi, P., & Van den Broeck, G. (2021, January). Probabilistic Sufficient Explanations. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*.

Wang, S., Minku, L. L., & Yao, X. (2018). A systematic study of online class imbalance learning with concept drift. *IEEE transactions on neural networks and learning systems*, *29*(10), 4802-4821.

Xu, R. (2021). Estimating Social Influence Effects in Networks Using A Latent Space Adjusted Approach in R. *The R Journal*.

You, X., Zhang, M., Ding, D., Feng, F., Huang, Y. (2021). Learning to learn the future: Modeling concept drifts in time series prediction. Proceedings of the 20th ACM International Conference on Information & Knowledge Management. https://doi.org/10.1145/3459637.3482271

Zhang, S., Tino, P., & Yao, X. (2021). Hierarchical reduced-space drift detection framework for multivariate supervised data streams. *IEEE Transactions on Knowledge and Data Engineering*.

McCallam, D., Braun, T., Akesson, B., Aspinall, D., Faganel, R., Guenther, H., ... & Varga, M.

(2021). *Final Report and Recommendations of the North Atlantic Treaty Organization (NATO) Research Task Group IST-129 on Predictive Analysis of Adversarial Cyber Behavior*. DEVCOM Army Research Laboratory US Naval Academy George Mason University Finnish Defence Research Agency University of Edinburgh Slovenia Ministry of Defence Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE) Defence Research and Development Canada US Naval Postgraduate School Estonian Business school Royal Military Academy Swedish Defence Research Agency Seetru Ltd Oxford University.

Gama, J., & Kosina, P. (2014). Concept drift in data streams: a review. ACM Computing Surveys (CSUR), 46(3), 1-37.

Kolter, J. Z., & Maloof, M. A. (2006). Dynamic Bayesian networks: Representation, inference and learning. Journal of Machine Learning Research, 7(Apr), 623-656.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sarkar, D. (2008). Lattice: Multivariate data visualization with R. Springer Science & Business Media.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

Zhang, Y., & Zhang, W. (2011). Predictive analytics in cyber security. Proceedings of the IEEE, 99(1), 156-173.

**16.2. Appendix A. Supplemental Material**

**Density Plots:**

****

**16.3. Appendix B. GitHub repository**

https://rb.gy/czen6z