RESEARCH ARTICLE OPEN ACCESS

# MARKET BASKET ANALYSIS

Mr.S.Sambasivam,M.C.A.,M.Phil.,[1],N.Nandhini[2],

[1]Professor, Department of Computer Applications ,

Nandha Engineering College (Autonomous), Erode, Tamilnadu,India.

[2]Final MCA, Department of Computer Applications,

Nandha Engineering College (Autonomous), Erode, Tamilnadu,India.

Email : [1]sammy2173@gmail.com, [2]nnandy914@gmail.com

**Abstract.** Market Basket Analysis or MBA is a field of modelling ways grounded upon the proposition that if you buy a certain group of particulars, you're further (or lower) likely to buy another group of particulars. MBA includes determination and vaticination client's geste grounded on expenditure pattern of former guests. MBA is applied not only for retail industries but also for a great number of other industries. There are studies which point to MBA and contribute to adding inflows in hospices operation by offering more seductive fresh services for new and regular guests. MBA grounded on multidimensional log it model was used to conduct a study Request handbasket analysis is to make a choice of purchasing, sailing or power of stocks in an equity request. Data booby-trapping ways insure high perfection of vaticination of stock price movement. In this thesis using MBA for perfecting styles of arranging products on store shelves was linked. Analysis of the most frequent guests' deals was performed. In this design, Request handbasket vaticination, i.e., supplying the client a shopping list for the coming purchase according to her current requirements, is one of these services. Current approaches aren't able of landing at the same time the different factors impacting the client's decision processco-occurrence, sequentuality, periodicity and recurrency of the bought particulars. To this end, this design defines a pattern Temporal Annotated Recurring Sequence (Seamen) suitable to capture contemporaneously and adaptively all these factors. We define the system to prize TARS and develop a predictor for coming handbasket named TBP (TARS Based Predictor) that, on top of TARS, is suitable to understand the position of the client's stocks and recommend the set of utmost necessary particulars. By espousing the TBP the supermarket chains could crop acclimatized suggestions for each individual client which in turn could effectively speed up their shopping sessions.

**Keywords:** Data Mining, Market Basket Analysis, Temporal Annotated Recurring Sequence.

## I. INTRODUCTION

Detecting purchase habits and their elaboration in time is a pivotal challenge for effective marketing programs and engagement strategies. In this environment, one of the most promising installations retail requests can offer to their guests is basket analysis, i.e., the automated soothsaying of the coming basket analysis that a client will buy. An effective basket recommender can act as a shopping list memorial suggesting the particulars that the client could presumably need.

A successful consummation of this operation requires an in- depth knowledge of an existent's shopping geste (1). The purchasing patterns of individualities evolve in time and can witness changes due to both environmental reasons, like seasonality of products or retail programs, and particular reasons, like diet changes or shift in particular preferences.

Therefore, a satisfactory result to basket analysis must be adaptive to the elaboration of a client's geste, the rush of her purchase patterns, and their periodic changes. This design proposes the Temporal Annotated Recurring Sequences (Seamen), adaptive patterns which model an existent's purchasing geste by four main characteristics.

First, TARS consider theco-occurrence a client totally purchases a set of particulars together. Secondly, Seamen model the sequentiality of purchases, i.e., the fact that a client totally purchases a set of particulars after another bone. Third, Seamen consider periodicity a client can totally make a successional purchase only in specific ages of the time, because of environmental factors or particular reasons. Fourth, TARS consider the recurrency of a successional purchase during each period, i.e., how constantly that successional purchase appears during a client's period of the time.

Modeling these four aspects –co-occurrence, sequentiality, periodicity and recurrency – is abecedarian to descry an existent's shopping geste and its elaboration in time. On one hand, unborn requirements depend on the requirements formerly satisfied what a client will buy depends on what she formerly bought.

On the other hand, the requirements of a client depend on her specific habits, i.e., recreating purchases she makes over and over. Far from being static, shopping habits are affected by both endogenous and particular factors. For this reason, periodicity is a pivotal specific of an adaptive model for basket analysis. This design exploits the Seamen to construct a parameter-free TARS Grounded Predictor (TBP) which solves the basket analysis problem and provides a basket recommendation as a list of particulars to be reminded in the coming purchase.

This design demonstrates the effectiveness of our approach by rooting the Seamen for thousands of guests in three large-scale real- world datasets. One of the main parcels of TARS is their interpretability, which allows retail chains to gain useful perceptivity about the guests' copping patterns. It shows that TARS can be used to infer important characteristics of products, like seasonality and inter-purchase times, which can be fluently interpreted by both a simple fine memorandum and a visual representation.

Also, we compare TBP with a force of state-of-the- art styles and show that

(i) TBP outperforms being styles,

(ii) TBP can prognosticate up to the coming 20 baskets,

(iii) the quality of TBP's prognostications stabilizes after about 36 weeks. Seamen and TBP are stoner-centric approaches given a client, they only use the client's individual data to prognosticate her unborn baskets. This aspect eases the guests' particular data operation and allows for developing acclimatized recommenders that can run on particular mobile bias.

It's also reviewed and distributed the affiliated work on transactional data mining for prognostications and recommendations. Coming basket analysis is an operation of recommender systems grounded on implicit feedback where only positive compliances (e.g., purchases or clicks) are available, and no unequivocal preferences (e.g., conditions) are expressed. The implicit feedback are given in a form of succession transactional data attained by tracking the druggies' geste over time, e.g. a retail store records the deals of guests through dedication cards.

Coming basket analysis is substantially aimed at the construction of effective recommender systems (or recommenders). Recommenders can be distributed into general, successional, pattern-grounded, and mongrel recommenders. General recommenders are grounded on cooperative filtering and produce recommendations for a client grounded on general guests' preferences. They don't consider any successional information ( i.e., which item is bought after which) and don't acclimatize to the guests' recent purchases. In discrepancy, successional recommenders are grounded on Markov chains and produce recommendations for a client exploiting successional information and recent purchase). Pattern grounded recommenders base prognostications on frequent itemsets uprooted from the purchase history of all guests while discarding successional information (21), (22),. Pattern grounded approaches constantly exploit or extend the Apriori algorithm for rooting the patterns.

## II. LITERATURE REVIEW

In this paper (1) the authors said that in the last times numerous accurate decision support systems have been constructed as black boxes, that's as systems that hide their internal sense to the stoner. This explanation lack constitutes both practical as well as ethical issue. The literature reports numerous approaches aimed at prostrating this pivotal weakness occasionally at the cost of panicking delicacy for interpretability. The operations in which black box decision systems are used colorful, and each approach is developed generally to yield a result for a specific problem and, as a consequence, delineating explicitly or implicitly its own description of interpretability and explanation.

The end of this paper is to give a bracket of the main problems addressed in the literature with respect to the notion of explanation and the type of black box system. Given a problem description, a black box type, and a asked explanation this check should help the experimenter to find the proffers more useful for his own work. The proposed bracket of approaches to open black box models should also be useful for putting the numerous exploration open questions in perspective.

Past ten years has witnessed the rise in opaque decision systems which are ubiquitous. These black box systems exploit sophisticated machine literacy models to prognosticate individual information that may also be sensitive. We can consider credit score, insurance threat, health status, as exemplifications. Machine literacy algorithms make prophetic models which are suitable to collude stoner features into a class ( outgrowth or decision) thanks to a literacy phase.

This literacy process is made possible by the digital traces that people leave behind them while performing everyday conditioning (e.g., movements, purchases, commentary in social networks,etc.). This enormous quantum of data may contain mortal impulses and prejudices. Therefore, decision models learned on them may inherit similar impulses, conceivably leading to illegal and wrong opinions.

The European Parliament lately espoused the General Data Protection Regulation (GDPR), which will come law in May 2018. An innovative aspect of the GDPR, which has been much batted, are the clauses on automated (algorithmic) individual decision- timber, including profiling, which for the first time introduce, to some extent, a right of explanation for all individualities to gain " meaningful explanations of the sense involved" when automated decision timber takes place.

Despite divergent opinions among legal scholars regarding the real compass of these clauses (6, 7, 15), everybody agrees that the need for the perpetration of such a principle is critical and that it represents moment a huge open scientific challenge. Without an enabling technology able of explaining the sense of black boxes, the right to an explanation will remain a " dead letter". By counting on sophisticated machine literacy models trained on massive datasets thanks to scalable, high- performance architectures, we risk to produce and use decision systems that we don't really understand. This impacts not only information on ethics, but also on safety and on artificial liability.

Companies decreasingly request services and products by bedding machine literacy factors, frequently in safety-critical diligence similar as tone- driving buses, robotic sidekicks, and substantiated drug. Another essential threat of these factors is the possibility of inadvertently making wrong opinions, learned from vestiges or spurious correlations in the training data, similar as feting an object in a picture by the parcels of the background or lighting, due to a methodical bias in training data collection.

How can companies trust their products without understanding and validating the underpinning explanation of their machine literacy factors? Gartner predicts that " by 2018 half of business ethics violations will do through the indecorous use of Big Data analytics". Explanation technologies are an immense help to companies for creating safer, more responsible products, and better managing any possible liability they may have. Likewise, the use of machine literacy models in scientific exploration, for illustration in drug, biology, socio-profitable lores, requires an explanation not only for trust and acceptance of results, but also for the sake of the openness of scientific discovery and the progress of exploration. As a consequence, explanation is at the heart of a responsible, open data wisdom, across multiple assiduity sectors and scientific disciplines.

Different scientific communities studied the problem of explaining machine literacy decision models. Still, each community addresses the problem from a different perspective and provides a different meaning to explanation. Utmost of the workshop in the literature come from the machine literacy and data mining communities.

The first one is substantially concentrated on describing how black boxes work, while the alternate bone is more interested in explaining the opinions indeed without understanding the details on how the opaque decision systems work in general. Despite the fact that interpretable machine literacy has been a content for quite some time and entered lately important attention, moment there are numerous ad-hoc scattered results, and a methodical association and bracket of these methodologies is missing.

Numerous questions feed the papers in the literature proposing methodologies for interpreting black box systems (9, 10) What does it mean that a model is interpretable or transparent? What's an explanation? When a model or an explanation is scrutable? Which is the stylish way to give an explanation and which kind of model is more interpretable? Which are the problems taking interpretable models/ prognostications? What kind of decision data is affected? Which type of data records is more scrutable? How important are we willing to lose in vaticination delicacy to gain any form of interpretability? We believe that a clear bracket considering contemporaneously all these aspects is demanded to organize the body of knowledge about exploration probing methodologies for opening and understanding the black box.

Being workshop tend to give just a general overview of the problem pressing unanswered questions and problems. On the other hand, other workshop concentrate on particular aspects like the impact of representation formats on comprehensibility (11), or the interpretability issues in term of advantages and disadvantages of named prophetic models (12). Accordingly, after feting four orders of problems and a set of ways to give an explanation, we've chosen to group the methodologies for opening and understanding black box predictors by considering contemporaneously the problem they're facing, the class of results proposed for the explanation, the kind of data anatomized and the type of predictor explained.

Need for Interpretable Models: Which are the real problems taking interpretable models and resolvable prognostications? In this section, they compactly report some cases showing how and why black boxes can be dangerous. Indeed, delegating opinions to black boxes without the possibility of an interpretation may be critical, can produce demarcation and trust issues. Training a classifier on literal datasets, reporting mortal opinions, could lead to the discovery of aboriginal prepossessions (13). Also, since these rules can be deeply concealed within the trained classifier, they risk considering, perhaps unconsciously, similar practices and prejudices as general rules. They're advised about a growing " black box society" (14), governed by " secret algorithms defended by artificial secretiveness, legal protections, obfuscation, so that purposeful or unintentional demarcation becomes unnoticeable and mitigation becomes insolvable.

They concluded that they've presented a comprehensive overview of styles proposed in the literature for explaining decision systems grounded on opaque and obscure machine literacy models. First, they've linked the different factors of the family of the explanation problems. In particular, we've handed a formal description of each problem belonging to that family landing for each bone the proper peculiarity. They've named these problems black box model explanation problem, black box outgrowth explanation problem, black box examination problem and transparent box design problem. Also, we've proposed a bracket of styles studied in the literature which takes into account the following confines the specific explanation problem addressed, the type of explanator espoused, the black box model opened, and the data type used as input by that black box model.

As shown in this paper, a considerable quantum of work has formerly been done in different scientific communities and especially in the machine literacy and data mining communities. The first one is substantially concentrated on describing how the black boxes work, while the alternate bone is more interested into explaining the opinions indeed without understanding details in how opaque decision systems work. The analysis of the literature conducted in this paper has led to the conclusion that despite numerous approaches have been proposed to explain black boxes, some important scientific questions still remain unanswered.

One of the most important open problems is that, until now, there's no agreement on what an explanation is. Indeed, some workshop give as explanation a set of rules, others a decision tree, others a prototype ( especially in the environment of images). It's apparent that the exploration exertion in this field fully ignored the significance of studying a general and common formalism for defining an explanation, relating which are the parcels that an explanation should guarantee,e.g., soundness, absoluteness, conciseness and comprehensibility.

Concerning this last property, there's no work that seriously addresses the problem of quantifying the grade of comprehensibility of an explanation for humans, although it's of abecedarian significance. The study of measures suitable to capture this aspect is grueling because it also consider also aspects like the moxie of the stoner or the quantum of time available to understand the explanation. The description of a ( fine) formalism for explanations and of tools for measuring how much an explanation is scrutable for humans would ameliorate the practical connection of utmost of the approaches presented in this paper.

Also, there are other open exploration questions related to black boxes and explanations that are starting to be treated by the scientific community and that earn attention and further disquisition. A common supposition of all orders of workshop presented in this paper is that the features used by the black box decision system are fully known.

Still, a black box might use fresh information besides that explicitly asked to the stoner. For illustration, it might link the stoner's information with different data sources for accelerating the data to be exploited for the vaticination. Thus, an important aspect to be delved is to understand how an explanation might also be deduced in cases where black box systems make opinions in presence of latent features. An intriguing starting point for this exploration direction is the frame proposed in (15) by Lakkaraju etal. for the evaluation of the vaticination models performances on labeled data where the decision of decision makers (either humans or black-boxes) is taken in the presence of unobserved features.

Another open exploration question is related to furnishing explanations in the field of recommender systems. When a suggestion is handed to a client, it should come together with the reasons for this recommendation. In (16) the authors define a case- grounded logic approach to induce recommendations with the occasion of carrying both the explanation of the recommendation process and of the produced recommendations. Incipiently, a further intriguing point is the fact that explanations are important on their own and predictors might be learned directly from explanations. A starting study of this aspect is (17) that presents a software agent learned to pretend the Mario Bros game only exercising explanations rather than the logs of former plays.

In this paper (2) the rise of smartphones and web services made possible the large-scale collection of particular metadata. Information about individualities' position, phone call logs, or web- quests, is collected and used intensely by associations and big data experimenters. Metadata has still yet to realize its full eventuality.

Sequestration and legal enterprises, as well as the lack of specialized results for particular metadata operation is precluding metadata from being participated and conformed under the control of the existent. This lack of access and control is likewise fueling growing enterprises, as it prevents individualities from understanding and managing the pitfalls associated with the collection and use of their data.

Their donation is two-fold

( 1) They described openPDS, a particular metadata operation frame that allows individualities to collect, store, and give finegrained access to their metadata to third parties. It has been enforced in two field studies;

(2) They introduced and dissect SafeAnswers, a new and practical way of guarding the sequestration of metadata at an individual position. SafeAnswers turns a hard anonymization problem into a further compliant security one. It allows services to ask questions whose answers are calculated against the metadata rather of trying to anonymize individualities'metadata.

The dimensionality of the data participated with the services is reduced from high-dimensional metadata to low-dimensional answers that are less likely to bere-identifiable and to contain sensitive information. These answers can also be directly participated collectively or in total. openPDS and SafeAnswers give a new way of stoutly guarding particular metadata, thereby supporting the creation of smart data- driven services and data wisdom exploration.

Particular metadata – digital information about druggies' position, phone call logs, or web- quests – is really the canvas of ultramodern data-ferocious wisdom and of the online frugality. This high-dimensional metadata is what allow apps to give smart services and substantiated gests. From Google's hunt to Netflix's ‘‘ pictures you should really watch, ’’ from Pandora to Amazon, metadata is used by marketable algorithms to help druggies come more connected, productive, and entertained. In wisdom, this high-dimensional metadata is formerly used to quantify the impact of mortal mobility on malaria or to study the link between social insulation and profitable development.

Metadata has still yet to realize its full eventuality. This data is presently collected and stored by hundreds of different services and companies. Similar fragmentation makes the metadata inapproachable to innovative services, experimenters, and frequently indeed to the existent who generated it in the first place. On the one hand, the lack of access and control of individualities over their metadata is fueling growing enterprises. This makes it veritably hard, if not insolvable, for an individual to understand and manage the associated pitfalls.

On the other hand, sequestration and legal enterprises are precluding metadata from being conformed and made astronomically accessible, substantially because of enterprises over the threat of reidentification.

Then they introduced openPDS, a field- tested particular data store (PDS) allowing druggies to collect, store, and give fine-granulated access to their metadata to third parties. We also introduce SafeAnswers, a new and practical way of guarding the sequestration of metadata through a question and answer system. Moving forward, advancements in using and booby-trapping these metadata have to evolve in resemblant with considerations of control and sequestration. openPDS and SafeAnswers allow particular metadata to be safely participated and conformed under the control of the existent.

Towards Personal Data Stores While questions of data power and the creation of depositories of particular data have been bandied for a long time, their deployment on a large-scale is a funk-and-egg problem; druggies are staying for compatible services while services are staying for stoner relinquishment. Exposures of the collection and use of metadata by governments and companies have still lately drawn attention to their eventuality. The combination of

1) A public interest in questions of control but also use of their data,

.2) Political and legal support on data power and

3) The scale at which metadata can now be collected and reused, might spark the large-scale deployment of PDS.

OpenPDS completely aligns with these trends. It uses the World Economic Forum description of ‘‘ power ’’ of metadata the rights of possession, use, and disposal. It follows programs of the National Strategy for Trust Individualities in Cyberspace (NSTIC) and explosively aligns with the European Commission’s reform of the data protection rules.

Eventually, it recognizes that druggies are interacting with multitudinous data sources on a diurnal base. Interoperability is therefore not enough to achieve data power or address sequestration enterprises. Rather, openPDS implements a secure space acting as a centralized position where the stoner’s metadata can live. openPDS can be installed on any garçon under the control of the existent ( particular garçon, virtual machine, etc) or can be handed as a service (SaaS by independent software merchandisers or operation service providers). This allows druggies to view and reason about their metadata and to manage fine-granulated data access.

From a profitable viewpoint, data power by the individual unnaturally changes the current eco-system. It enables a fair and effective request for metadata – a request where druggies can get the stylish services and algorithms for their metadata. Druggies can decide whether a service provides enough value for the quantum of data it requests, and services can be rated and estimated. Druggies are empowered to ask questions like “is chancing out the name of this song worth enough to me to give away my position?” Druggies can seamlessly give new services access to their history and present metadata while retaining power.

From a business viewpoint, similar data power is likely to help foster druthers to the current data-selling and advertising- grounded business model. New business models fastening on furnishing tackle for data collection, storehouse for metadata, or algorithms for better using metadata might crop while software for data collection and data operation might be substantially open- source. The proposed frame removes walls to entry for new businesses, allowing the most innovative algorithmic companies to give better data-powered services.

Other approaches have been proposed for the storehouse, access control, and sequestration of data. Former approaches fall into two orders pall storehouse systems and particular data depositories. First, pall storehouse systems, similar as the bones that have been commercially developed by companies like Dropbox and Carbonite, are a first approximation of a stoner- controlled information depository for particular data. They still concentrate on storing lines and only apply the most introductory type of access control, generally on a train or brochure base. They don't suggest any data aggregation mechanisms and, formerly access has been granted, the raw data is exposed to the external world, potentially compromising sequestration.

Second, particular data depositories have been developed in academic and marketable settings (18 – 20). All of these depositories are still confined to specific queries on a particular type of data, similar as interests or social security figures. They give only a introductory access- control position, which means that formerly access to the data is authorized, sequestration may be compromised. openPDS differs from former approaches in its alignment with current political and legal thinking, its focus on large-scale metadata, and its SafeAnswers sequestration- conserving medium.

They concluded that eventually, as technologists and scientists, we're convinced that there's an amazing eventuality in particular metadata, but also that benefits should be balanced with pitfalls. By reducing the dimensionality of the metadata on-the- cover and laboriously guarding druggies, openPDS/ SafeAnswers opens up a new way for individualities to recapture control over their sequestration. openPDS/ SafeAnswers still still face a number of challenges.

Each challenge includes several implicit directions for unborn exploration

(1) The automatic or semi-automatic confirmation of the processing done by a PDS module;

(2) The development of SafeAnswers sequestration-conserving ways at an individual position for high-dimensional and ever- evolving data (mobility data, accelerometer readings, etc.) grounded on being anomaly discovery frame and potentially stored in largely-decentralized systems;

(3) The development or adaption of sequestration conserving datamining algorithms to an ecosystem conforming of distributed PDSs; and

( 4) UIs allowing the stoner to more understand the pitfalls associated with large-scale metadata and to cover and fantasize the metadata used by operations.

## III. PROPOSED METHODOLOGY

The market basket method scheme allows adapting external non-homogeneous data sources for fitting into previous recommendation systems. The integration scheme is based on several market basket analysis methods, such as association rule s, collaborative filtering and clusterization. Two datasets are taken with transaction of supermarkets' customers data for the experimental study. The existing model defines two metrics to evaluate the quality of constructed recommendation system and conduct experiments comparing the original and adapted recommendation systems.

The output shows the ability for improving recommendation system quality by using additional non-homogeneous data sources. But, the datasets are extremely varying in nature and do not allow to get the expected results. The study's main result is that we understood further ways of the research that it is required to find more suitable datasets and improve the used methods and mode. Drawbacks are:

1. Only Item–to-item collaborative filtering and recommendations are generated based on a similarity of clients transaction details.
2. Improve the quality of constructed RS by using additional data sources
3. Common groups of rules by using clusterization methods and filtering of non-conforming parts of data.
4. Do not find missing date of customer Transaction data
5. Does not identifying missing item set with last transaction data.
6. Do not predicated missing item with loss of profit by customer purchase transaction dataset.

The Temporal Annotated Recurring Sequences adaptive patterns which model an existent's purchasing geste by four main characteristics. First, TARS consider theco-occurrence a client totally purchases a set of particulars together. Secondly, Seamen model the sequentiality of purchases, i.e., the fact that a client totally purchases a set of particulars after another bone. Third, Seamen consider periodicity a client can totally make a successional purchase only in specific ages of the time, because of environmental factors or particular reasons. Fourth, TARS consider the recurrency of a successional purchase during each period, i.e., how constantly that successional purchase appears during a client's period of the time. Modeling these four aspects –co-occurrence sequentiality, periodicity and recurrency – is abecedarian to descry an existent's shopping geste and its elaboration in time.

## IV. FINDINGS

• The system can prognosticate up to 20 baskets. The quality of analysis stabilizes after about 36 weeks. Seamen and TBP are stoner-centric approaches given a client, they only use the client's individual data to prognosticate her unborn baskets.

• This aspect eases the guests' particular data operation and allows for developing acclimatized recommenders that can run on particular mobile bias.

• Finds missing date of client Sale data

• Relating missing item set with last sale data.

• Rested missing item with loss of profit by client purchase sale dataset.

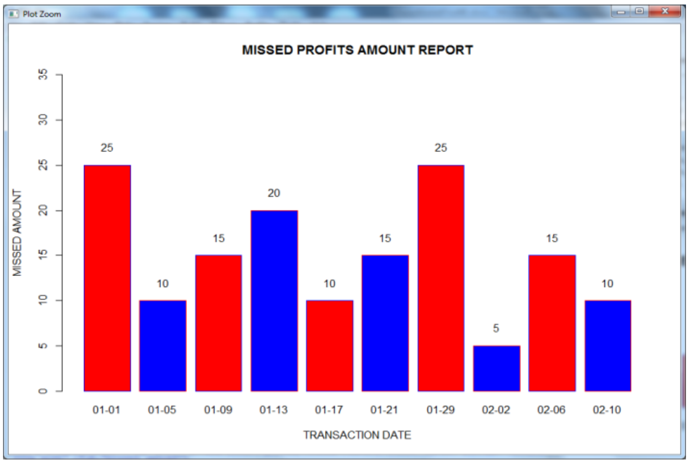


FIG 4.1 MISSING PROFITS AMOUNT REPORT

## V. CONCLUSION

In this work, we've proposed a data-driven, interpretable and stoner-centric approach for request handbasket vaticination. We've defined Temporal Annotated Recurring Sequences and used them to construct a TARS Based Prediction (TBP) for coming handbasket soothsaying. Being parameter-free,

TBP leverages the particularity of the guests' geste to acclimate the way Seamen are uprooted, therefore producing further substantiated patterns. We've performed trials on real- world datasets showing that TBP outperforms state-of-the- art styles. Inversely important, we've shown that the birth of Seamen provides precious interpretable patterns that can be used to gather perceptivity on both the guests' Copping actions and products' parcels like season- ality andinter-purchase times. Our results demonstrate that at least 36 weeks of a client's purchase geste are demanded to effectively prognosticate her unborn baskets. In this Script, TBP can effectively prognosticate the posterior twenty unborn baskets with remarkable delicacy.

REFERENCES


[1] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini et al., "A survey of methods for explaining black box models," ACM Computing Surveys (CSUR), vol. 51, no. 5, p. 93, 2018.

[2] Y.-A. de Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland, "openpds: Protecting the privacy of metadata through safeanswers," PloS one, vol. 9, no. 7, p. e98790, 2014.

[3] M. Vescovi, C. Perentis, C. Leonardi, B. Lepri, and C. Moiso, "My data store: toward user awareness and control on personal data," in Ubicomp. ACM, 2014, pp. 179–182.

[4] K. Christidis et al., "Exploring customer preferences with probabilistic topic models," in ECML-PKDD, 2010.

[5] Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-ofthe-art and possible extensions, Knowledge and Data Engineering, IEEE Transactions, vol. 17, 734—749 (2005).

[6] B. Goodman and S. Flaxman. Eu regulations on algorithmic decision-making and a right to explanation. In ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. http://arxiv. org/abs/1606.08813 v1, 2016.

[7] S. Wachter, B. Mittelstadt, and L. Floridi. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2):76–99, 2017.

[8] G. Comand`e. Regulating algorithms regulation? first ethico-legal principles, problems, and opportunities of algorithms. In Transparent Data Mining for Big and Small Data, pages 169–206. Springer, 2017.

[9] A. Weller. Challenges for transparency. arXiv preprint arXiv:1708.01870, 2017.

[10] J. M. Hofman, A. Sharma, and D. J. Watts. Prediction and explanation in social systems. Science, 355(6324):486–488, 2017.

[11] J. Huysmans, K. Dejaeger, C. Mues, J. Vanthienen, and B. Baesens. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. Decision Support Systems, 51(1):141–154, 2011.

[12] A. A. Freitas. Comprehensible classification models: a position paper. ACM SIGKDD explorations newsletter, 15(1):1–10, 2014.

[13] D. Pedreshi, S. Ruggieri, and F. Turini. Discrimination-aware data mining. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 560–568. ACM, 2008.

[14] F. Pasquale. The black box society: The secret algorithms that control money and information. Harvard University Press, 2015.

[15] H. Lakkaraju, J. Kleinberg, J. Leskovec, J. Ludwig, and S. Mullainathan. The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 275–284. ACM, 2017.

[16] D. McSherry. Explanation in recommender systems. Artificial Intelligence Review, 24(2):179–197, 2005.

[17] S. Krening, B. Harrison, K. M. Feigh, C. L. Isbell, M. Riedl, and A. Thomaz. Learning from explanations using sentiment and advice in rl. IEEE Transactions on Cognitive and Developmental Systems, 9(1):44–55, 2017.