Financial Services Data and Analytics Newsletter

August 2022

pwc

Click to launch ▶

Introduction

To meet the information needs of various stakeholders across business units, technology and data engineering teams generate various insights from data analysis.

Statistical models or supervised learning models have been used in the financial services industry for decades now. Such a model involves forming a hypothesis first and then proving or disproving it using available data. The efficacy of statistical models can be impacted over time – primarily due to inefficient variables, a shift in the hypothesis, emerging unknown patterns, incorrect or incomplete labelling of data, etc. Corrective actions should hence be taken periodically to re-calibrate the model and verify its robustness. But what if one is unable to create a meaningful hypothesis because of an incomplete understanding of the data?

In this edition's topic of the month, we take a look at how unsupervised learning can help in solving the above problem. In the case of an unsupervised learning model, one does not need to build a hypothesis upfront, but rather analyse the available data, establish relationships between the data points and gain new insights into unknown patterns. Data is parsed and labelled and unknown scenarios develop into 'now-known' ones, which helps to uncover or formulate a solid hypothesis and then build future capabilities for supervised learning models.

The newsletter also includes updates on transformational changes proposed by regulatory bodies such as the RBI and IRDA to protect consumer interests and respond to changes in industry dynamics. Happy reading!

Topic of the month: Unsupervised machine learning

1. Machine learning (ML) and its types

ML enables decision-making intelligence by training algorithms to learn from data patterns. Simply put, in ML, users provide input data to the computer algorithms, which analyse it for patterns and arrive at decisions/recommendations based on identified data patterns.

ML can be divided into three types:

- supervised learning models
- unsupervised learning models
- reinforcement learning models.

Supervised ML models are algorithms that are trained using a labelled or tagged dataset, i.e. data with a known set of variables that are classified as independent (predictors) and labelled/dependent (target). While supervised ML algorithms are heavily dependent on labels, absence of labelled data can also be used for decision making.

Unsupervised ML algorithms help us derive insights from data sets that do not have a target/labelled variable.

Reinforcement learning, as the name suggests, is when the algorithm learns by itself through trial and error. It is a feedback-driven process that learns from experience and improves performance.

2. When to use unsupervised ML

Unsupervised ML models are independent learning algorithms that explore unknown scenarios/patterns or data groupings without any supervision or guidance. Lack of supervision or human intervention in these models also eliminates human bias that may creep in while working with labelled data.

Unsupervised ML algorithms can be leveraged in the following situations:

- lack of labelled data
- need to recognise intuitive patterns from data without providing prior learning to the algorithm
- need to group or cluster data based on similarities with no prior knowledge of how many groups or clusters should exist
- detecting and flagging edge cases or patterns in data
- generalisation of the model due to changing data patterns.

Topic of the month: Unsupervised machine learning

3. Applications of unsupervised ML algorithms:

Unsupervised ML identifies homogeneous and heterogeneous patterns in information provided. Applications of these algorithms are briefly described below:

- **Clustering/segmentation:** A distance-based technique is used to segment or cluster data points based on inherent homogeneity or similarity. This is typically used to identify cohorts of customers, consumers or groups that display similarities in behavioural patterns. Customers, consumers or groups thus identified are then targeted with distinct strategies with respect to products/services being offered to them. For example, in insurance, we can identify claims that are of a similar nature and then allocate adjusters or surveyors with a degree of specialisation, thus providing a hyper-personalised experience to customers.

  Application in the banking, financial services and insurance (BFSI) sector:

  a. Customer segmentation to identify best marketing segments for campaigns

  b. Tracking of customer behavioural patterns and profitability

  c. Clustering based on transaction type

  d. Product profiling and customer propensity to buy profiled products

  e. Grouping of policyholders according to risk or profitability

  f. Clustering based on types of claims

- **Dimension reduction:** This technique helps reduce the number of features used in the model by transforming them into key components or variables while retaining their characteristics. It is generally used in cases involving too many features or variables and where it is difficult to perform variable reduction without compromising on information loss. Principal component analysis (PCA) or independent component analysis (ICA) can then help to capture as much variability (information) in data as possible and to group the features through orthogonal components, thereby reducing the number of features required for model development.

  Application in BFSI:

  a. Transaction monitoring to detect fraud and compliance breaks

  b. Claims and underwriting fraud analytics

  c. Anti-money laundering (AML)

Topic of the month: Unsupervised machine learning

- **Anomaly detection:** Unsupervised ML can be used to identify or flag data points or outliers that are significantly different from the norm. For example, it can flag suspicious credit card transactions using algorithms like isolation forest based on a customer's past spending patterns and how a particular transaction differs from past ones.

  Application in BFSI:

  a. PCA use cases in image compression and computer vision

  b. Features engineering and reduction in high-dimensional datasets

- **Association rules:** This involves data-driven pattern recognition techniques (e.g. people buying X will also buy Y). For example, algorithms can be used to identify patterns or relationships between data points using apriorism rules such as a customer opting for internet banking would also opt for a debit card.

  Application in BFSI:

  a. Next best action – cross-selling bank products

  b. Recommendations on insurance policy add-ons

  c. Identifying associations in insurance claim invoices, e.g. vehicle part combinations in motor claims, disease-treatment-medicine combinations in health claims

  d. Identifying likelihood of investing in 'x' funds or investments if holding a portfolio in 'y' funds or investments
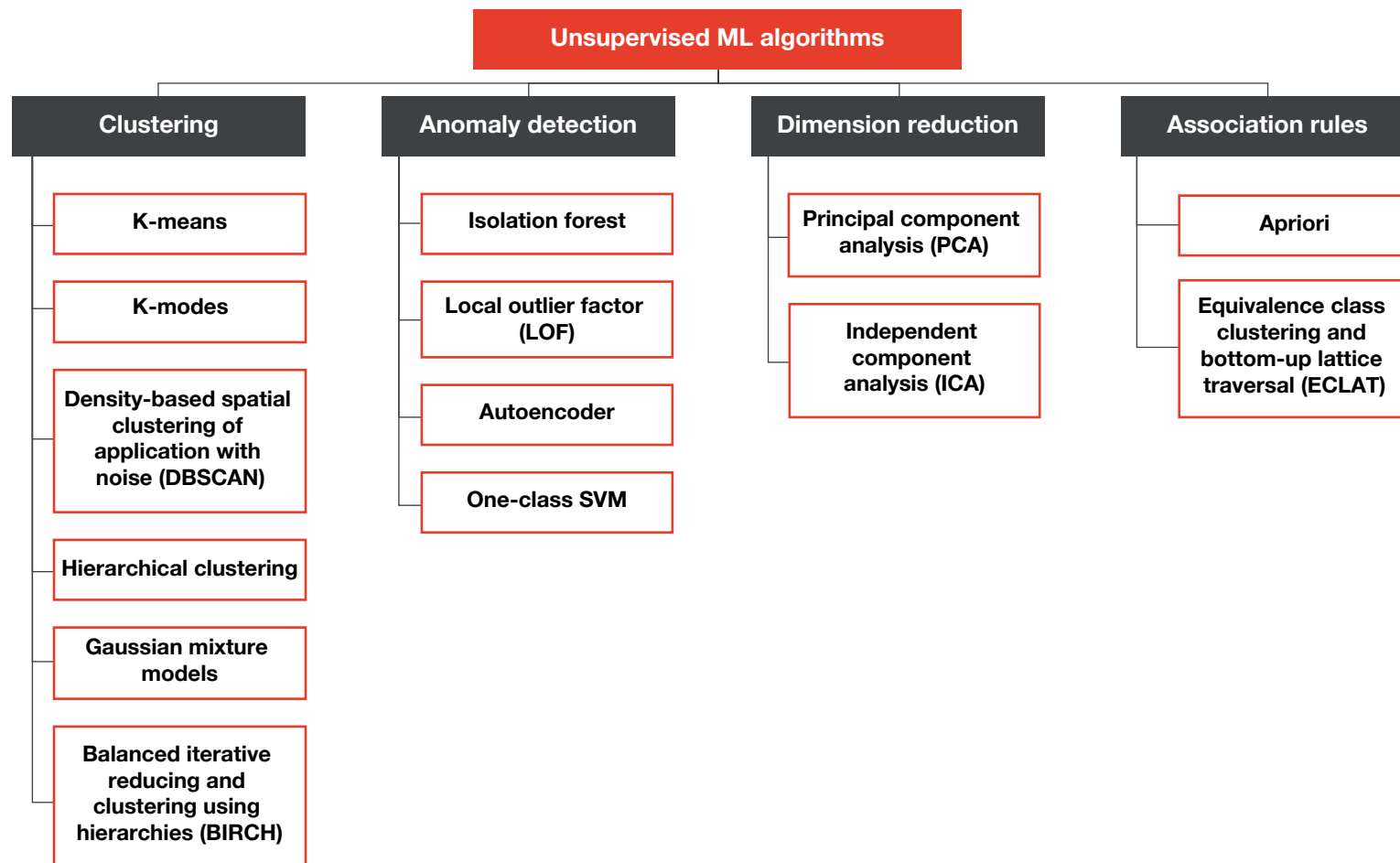
Topic of the month: Unsupervised machine learning

The figure below lists the commonly used unsupervised algorithms by application:

Unsupervised ML algorithms

Clustering	Anomaly detection	Dimension reduction	Association rules
K-means	Isolation forest	Principal component analysis (PCA)	Apriori
K-modes	Local outlier factor (LOF)	Independent component analysis (ICA)	Equivalence class clustering and bottom-up lattice traversal (ECLAT)
Density-based spatial clustering of application with noise (DBSCAN)	Autoencoder		
Hierarchical clustering	One-class SVM		
Gaussian mixture models			
Balanced iterative reducing and clustering using hierarchies (BIRCH)			

Source: PwC analysis

Topic of the month: Unsupervised machine learning

4. How unsupervised models fit into the ML landscape:

Unsupervised models can be leveraged to build capabilities for supervised models by solving for the latter's reliance on better labelled data. Unsupervised models offer the following benefits:

- **Better understanding of data dynamics:** Unsupervised ML algorithms like clustering or dimension-reduction techniques are often used as plug-ins before fitting a supervised model in order to better understand the underlying data dynamics, reduce redundancies in data and achieve better accuracy.

- **Identifying and converting 'unknown-unknown' risks into 'now-known' risk scenarios:** Unknown risks and patterns are not accounted for in supervised ML algorithms as they learn and train for known patterns tracked historically. Here, unsupervised anomaly detection algorithms play a key role in flagging unknown risk patterns by holistically evaluating the data points and identifying edge cases. These unknown risk patterns could be validated and streamlined or converted into 'now-known' patterns and tracked pre-emptively. This helps in developing a flexible model that is capable of dealing with new and changing data patterns.

- **Enriching data labels:** Unsupervised models could ultimately contribute towards the enrichment of data. These models will constantly look for patterns in data by grouping them according to similarity, thus adding new dimensions for data segregation or tagging and labelling suspicious trends. Tracking these labels over time would be cost-effective and help build a good training dataset for supervised ML algorithms with better learning nudges.

PwC's implementation experience:

Problem statement: Transaction monitoring for a firm's general ledger (GL) transactions across products to identify and pre-emptively track unknown risk or transaction patterns

Solution: PwC has implemented its Anomaly Detection Platform for transaction monitoring. The platform identifies unknown risks or transaction patterns in data by leveraging an unsupervised algorithm – isolation forest – to flag anomalies. These flagged labels are then fed to supervised decision trees to build explainability and detect patterns for validation. Validated patterns are streamlined into the current monitoring process. The solution is flexible across data trends, and the feedback tracked over time for these patterns will be leveraged to build a robust supervised learning engine to track anomalies, thus reducing complexity and run time.

Topic of the month: Unsupervised machine learning

Impact: The solution expands the user's outlook on identifying possible risk patterns – for example, flagging accounts with **sudden spikes in transaction amounts, seldom or rarely seen types of transactions within accounts, incorrect GL tagging for type of transaction or incorrect currency tagging for transactions, and defining new thresholds for certain types of transactions.** It automatically flags suspicious and out-of-the-norm trends in the data. The insights provided in terms of deterministic rules or patterns make it easier for the user to plug in these checks into the monitoring process and account for any unknown leakages. Thus, it helps in building a robust monitoring process.

5. Challenges of unsupervised ML

While unsupervised ML algorithms are a great choice when availability of data labels is an issue, one has to address some challenges that accompany their:

- Computational complexity: Algorithms can be trained on the fly; however, processing high volumes of data increases training time and computation complexity.

- Validation: Patterns detected by the algorithms need to be validated manually and evaluated to determine if the model output is useful.

- Lack of supporting literature and research: Applications of unsupervised ML are not as well researched or widely available as in the case of supervised ML models.

Industry news

1. Experian report says data and analytics top priority for lenders in short term

A survey of 164 decision makers from the financial services sectors in India, Australia and Indonesia found that 84% wish to adopt AI urgently for credit risk analysis and 67% expect real-time data and analytics for investments. Further, 66% feel legacy system dependency inhibits automation, while 36% feel data standardisation is a barrier to automation in credit risk management

2. Decentraland launches the first metaverse ATM

The world's first metaverse ATM was launched on the Decentraland platform in partnership with the Metaverse Architects studio and the payment gateway Transak. The ATM will make Web 3.0 transactions similar to real-world ATM transactions where users can purchase MANA cryptocurrency with fiat currency or any other cryptocurrency. The team believes the solution will increase conversion rates in metaverse stores and make it easier for users to perform transactions in Decentraland.

3. Indian banks reported 248 data breaches in the last 4 years

Indian banks have reported 248 successful data breaches in the last 4 years, mostly pertaining to credit card data breaches. Out of the 248 breaches, 41 were reported by public sector banks and the rest by private sector banks. On this matter, the Reserve Bank of India (RBI) has informed the Centre that it has issued cyber security guidelines for better controls on data with banks. The RBI has also asked banks to strengthen their IT risk governance framework and advised management to initiate actions against erring staff within a specific time period. As per the Ministry of Home Affairs, a total of 1.4 lakh cyber security incidents were recorded in 2021.

Knowledge bytes

1. **RBI governor says the future of Indian banking will depend on social media data and new technologies**

   The RBI governor discussed how banks can quickly offer personalised offerings to meet customer expectations using social media and analytics. Social media data offers the potential for enhanced customer acquisition, customer segmentation, financial inclusion and grievance management. He also emphasised the need for traditional banks to adopt technology or collaborate with FinTechs.

2. **Insurance Regulatory and Development Authority (IRDAI) is facilitating insurance business through multiple reforms**

   The IRDAI has developed a mechanism to speed up the registration process for new insurers and acquire certificate of registration. Additionally, to make the Pradhan Mantri Jeevan Jyoti Bima Yojana more accessible, the IRDAI has relaxed capital requirements. Greater flexibility will be provided to corporate agency tie-ups. While earlier each corporate agent could tie up with only three insurers each from the life, health and general insurance sector, this has now increased to nine for each sector. The regulator has also allowed insurance marketing firms to tie up with six insurance companies from each sector instead of the current cap of two. It has also recently released draft regulations and is seeking feedback from companies on proposing a 20% cap on insurance agent's commission and a limit on expenses-on-management (EoM) at 30% of gross premium for general and standalone health insurance companies in India.

3. **New set of digital lending guidelines by the RBI aimed at improving data security, increasing transparency and protecting customers' interests**

   In a press release dated 10 August 2022, the RBI issued the recommendations of the working group on digital lending implementation. The RBI has mandated that lenders refrain from accessing mobile phone resources such as files and media, contact lists, call logs and telephony functions. With clear consent from users, one-time access could be taken for necessary facilities.

4. **RBI might permit small finance banks (SFBs) to co-lend with non-banking financial companies (NBFCs)**

   Currently, SFBs are allowed to only lend directly. Co-lending with NBFCs will entail increased priority sector lending. Industry experts envision expertise-based lending – for instance, SFBs targeting the agriculture industry will co-lend with NBFCs currently operating in that field.
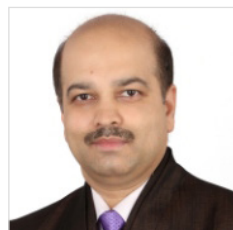
Contact us

Mukesh Deshpande

Partner, Technology Consulting
mukesh.deshpande@pwc.com
+91 98450 95391

Hetal Shah

Partner, Technology Consulting
hetal.d.shah@pwc.com
+91 9820025902

Acknowledgements

About PwC

At PwC, our purpose is to build trust in society and solve important problems. We're a network of firms in 155 countries with over 327,000 people who are committed to delivering quality in assurance, advisory and tax services. Find out more and tell us what matters to you by visiting us at www.pwc.com.

PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

pwc.in