Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 6s (2025)

Identification of Actionable Insights from Online Data Sources by using the Hashed Coherence Frequency Calculator (HCFC) Algorithm

Dr. B. Srinivasan

Associate Professor of Computer Science, School of Arts and Science, Vinayaka Mission's Chennai Campus, Vinayaka Mission's Research Foundation Deemed to be University, Salem, Paiyanoor, Tamil Nadu, India. Email: srinivasan.avca014a@avsas.ac.in

Article History:

Received: *16-10-2024*

Revised: *30-11-2024*

Accepted: *10-12-2024*

Abstract:

The present contemporary world is propelled by data. Everyone in society makes decisions based on the existing data and proceeds accordingly. Furthermore, all machinery is automated using advanced technologies such as Artificial Intelligence (AI), Internet of Things (IoT), Machine Learning (ML), and Data Science (DL), which predominantly harness data-driven insights for daily operations. Business organizations, social media, healthcare, big data analytics, and locality mapping are just a few modern industries that rely heavily on data-driven strategies. These areas utilize Cloud Computing (CC) to store the data in the cloud and allow access to data on any device. The primary consideration in data-driven terminology is that not all available data is utilized for decision-making. Hence, before adopting a data-driven process, it is essential to identify a collection of trustworthy and actionable insights that will enable better decision-making. This paper proposes a novel algorithm, the Hashed Coherence Frequency Calculator (HCF), for segregating actionable and non-actionable insights from the given dataset and enabling better decision-making. The algorithm is applied over two case studies containing Amazon product reviews and Google Play Store Apps review datasets. The coherent frequency count statistical measure is mainly applied to identify actionable insights, and the outcomes will also be compared with the existing approaches.

Keywords: actionable insights, cloud computing, coherence frequency count, topic detection, data-driven decision-making, Hash vectorizer, non-actionable insights, TF-IDF, LDA, outliers.

I. INTRODUCTION

Data has been used for decision-making from the dawn of time. Since the inception of information technology till the present day, data is the foundational for decision-making across all domains. Recent advancements in artificial intelligence and the advent of big data analytics and data science have propelled the present contemporary world to be powered by data. The modern world comprises diverse data, including corporate, banking, medical, social media, and data from various devices like the Internet of Things [1].

The significance of data-driven decision-making is escalating across numerous recent real-world applications. Websites like Amazon and Flipkart utilize client purchasing behaviour to provide tailored suggestions. Google Maps utilizes collected data to provide users with real-time updates on traffic situations. Moreover, healthcare applications offer insights into patients' states and suggest suitable treatments for particular disorders. Businesses utilize user input from social media to guide choices about product manufacture and sales expansion. All the domains above employ the cloud computing

infrastructure to manage data centrally. The core premise of data-driven terminology is that not all accessible data are employed in decision-making. Hence, it is essential to identify actionable insights that promote better decision-making before implementing data-driven concepts. Actionable insights are pertinent and instructive for a specific event or individual [2].

In the dynamic realm of present-day business, data utilization has become a pivotal predictor of success. Consequently, most firms are adopting data analytics as a strategic need to uncover meaningful insights, facilitate informed decision-making, and achieve a competitive advantage in their respective sectors [3].

Actionable insights foster the next level of ideas and create openings for breakthroughs that improve lives in all areas.

Finding of insights from the given dataset is an epiphany act and entails the identification of anomalies within the provided data. The identification of actionable insights entails the following

- Data alignment with objectives
- Selection of the appropriate instrument for the data analytics process.
- Emphasizing a human-centered approach, the user needs, and behaviors for ideal insight extraction.
- Transformation of actionable findings into actionable tactics.
- Utilization of the Human-Centered Insights Certificate to enhance the purification process.

Actionable insights must also be evaluated within the context of Personally Identifiable Information since they signify a specific category of products, services, and occasionally an individual. Given that the cloud computing environment is plagued by security and privacy concerns, appropriate security and privacy mitigation strategies must be implemented across all identified insights. This present research paper discusses a unique methodology for uncovering actionable insights using a newly constructed framework known as the Hashed Coherence Frequency Calculator(HCFC) algorithm, with the results assessed via two case studies dealing with Amazon product reviews and Google Play Store App reviews.

The opening section of this paper presents the essential backdrops, and the subsequent section examines the evaluation of the relevant literature. The supportive methods like coherence frequency count, Hashing Vectorizer, and Perplexity are discussed in the third section, followed by the overall proposed methodology framed in the fourth section. All the observed results with the sample datasets are analyzed in the results and analysis section, and the final section addresses the conclusion and future directions.

II. RELATED WORKS

The growing nature of data underscores the necessity of actionable insights to inform business decisions. Actionable insights are specific, data-driven recommendations that can be executed to move forward. Finding the needed information for making decisions can sometimes be challenging, even with the huge amount of available data. Large amounts of data may provide a wealth of ideas and information for creating new goods, services, systems, procedures, and experiences. A detailed understanding of converting raw data into meaningful and useful insights is essential for anyone engaged in data-driven decision-making. This literature review explores the diverse methodologies employed by various writers to extract actionable insights from the available data, along with their recommendations.

Nivedhaa N conducted a thorough study elucidating the progression from raw data to useful insights in data science [4]. Richard McCreadie and his team identified a lack of actionable insights for locating emergency responders in catastrophic events. They proposed a filtering method to extract usable insights and categorize requests during crises [5]. Jansen Bernad and the team investigated the relationship between avatars and analytics to demonstrate the efficacy of data-driven decision-making [6]. Alex Bogatu et al. developed schema and instance-based features to detect relevant patterns using hashtags [7]. A related study utilized data analytics to guide business decisions, enhance operational efficiency, and improve customer experiences [8]. Harish Narne investigated the groundbreaking potential of AI-driven data analytics in turning huge data into meaningful insights [9]. Arisa Shollo and Robert D. Galliers demonstrated how understanding, context, and subjectivity help organizations to create actionable insights. The authors illustrated that actionable insights represent an ongoing human endeavor and contemplated the consequences for future study [10]. Kristof Coussement and Dries F. Benoit outlined the principles of interpretable data science, which distill domain relationships to generate informed insights for decision-making. They also contextualized the current significance of interpretable data science in enhancing decision-making processes [11].

Ayuns Luz and Godwin Olaoye found that actionable data and effective engagement tactics may boost employee engagement, contentment, health, and corporate performance[12]. Gabelaia Ioseb used natural language processing, machine learning, and network analysis to unleash social media data. The author additionally found that such methods can help businesses, marketers, and decision-makers transform social media data into valuable insights [13]. Priya Patel and Gahletia Sumit examined how big data analytics might improve customer experiences by providing actionable information across individual touchpoints of the consumer journey [14].A. S. Rao, B. V. Vardhan, and H. Shaik examined Exploratory Data Analysis (EDA) to uncover insights and articulated its significance in the field of data analysis. The authors also examined EDA's related tools and packages [15]. J Hullman and A Gelman developed a Bayesian model to consolidate several analytical processes, advocating for a more deliberate integration of graphical inference methods employed in data visualization [16]. MK Manju, AO Philip, and MU Sreeja concentrated on doing EDA on the Indian Premier League (IPL) dataset, which encompasses historical match information to uncover latent insights and utilized these features for predicting match results [17].

O Badmus, SA Rajput, JB Arogundade, and M Williams elucidated the amalgamation of AI with Business Intelligence (BI) to provide profound insights from intricate datasets in real time [18]. P Rajan laid a thorough framework for marketing intelligence and IoT data analytics to support data-driven decision-making. The author also detailed how companies might get valuable insights [19]. Y Xu delineated a disclosure utilizing machine learning techniques to summarize client comments autonomously and to get insights from these summaries through Natural Language Processing (NLP) [20]. S. Al. Mesmari examined the concept of AI and cognitive computing in decision-making, highlighting how these tools may transform useless data into knowledge and outlining the benefits of applying them [21]. Second-order explainable AI (SOXAI) has been examined by EZ Zeng, H Gunraj, S Fernandez, and A Wong to acquire actionable insights that can be employed to improve the model's performance [22]. Kharakhash underscored the significance of data visualization tools for converting complicated data into usable insights. The author also noted that data visualization tools provide better

data comprehension, enable pattern discovery, and empower decision-makers to develop relevant conclusions [23].

O Olaniyi et al. examined the revolutionary potential of predictive analysis for extracting useful insights and decision-making, as well as the difficulties in drawing actionable conclusions from an unprecedented inflow of data [24] .E Hasan, M Rahman, C Ding, JX Huang, and S Raza performed a thorough investigation on review-based recommender systems. They elucidated many filtering and classification methodologies predicated on feature extraction and review strategies [25]. Y Yang and R Subramanyam developed a method termed stable LDA aimed at generating consistent outcomes in topic modeling, thereby facilitating the extraction of valuable insights [26].

The described literature review reveals that each work emphasized the need for usable insights from various fields. The evaluated experiments show how well various approaches can extract insightful values from intricate data sets, including machine learning, network analysis, exploratory data analysis, data visualization, and stable LDA topic detection. This research focuses on coherent-based topic modeling to extract actionable insights from the provided data.

III. METHODS USED FOR INSIGHTS DETECTION

The main objective of this research work is to uncover actionable insights within the dataset and to segregate outliers or non-actionable elements. This research focuses on text data to extract actionable insights and address outliers by utilizing the coherence frequency of word count. Coherence is a method used to assess the quality of topics in topic modeling by quantifying the semantic connection of words within a topic.

The following paragraphs demonstrate the base methods, which act as the key elements for constructing a novel approach to finding actionable insights in the present research work.

1) Coherence

Coherence is a metric that measures a subject modeling technique's interpretability by humans. It is determined by comparing two words lexically [28]. Coherence is often calculated as the sum of the correlations among the top n words that proceed from W_1. W_n is utilized to define a topic [29]. Coherence is calculated by using the following equation (1).

$$Coherence = \sum_{i<j} correlation(Wi, Wj) \qquad (1)$$

where W_i and W_j are the words, i and j are the word counts and should be less than the number of words n in the given data source, i.e., $i<j<n$.

In the context of uncovering actionable insights, the coherence frequency measure serves to pinpoint

1. Most significant topics: Examining the coherence frequency of words reveals the most significant topics and themes within a given text or corpus.
2. Recognizing sentiment patterns: Coherence frequency can assist in recognizing patterns in sentiment or emotional tone, including positive or negative feelings towards a specific topic.
3. Define entity relationships: The coherence word frequency count for linking a set of entities, such as social communities, issues on posts, and landmarks, is used to define the relationship among the

entities.

4. Insightful words: Coherence frequency helps to identify words or sentences that are very close or related to a particular topic.

2) Hashing Vectorizer (HV)

Hashing Vectorizer (HV) is a Natural Language Processing (NLP) technique used for transforming text data into numerical matrix format. HV helps the machine learning algorithms to process text data by means of tokenization and then followed by hashing, which maps tokens into numerical indices. facilitating the independent processing of each word [30]. HV is calculated using the equation provided in (2).

$$h(x) = (hash(x) mod\ n) \qquad (2)$$

where $x \in R^d$ is the input vector, $hash(x) \in Z$ is the hash value of x, $n \in Z$ is the number of features, and $h(x) \in \{0, 1, ..., n-1\}$ is the hashed index value.

Coherence and the HV collectively identify actionable insights from the given datasets. First, the hashing vectorizer is calculated by transforming each text data point into numerical values. Then, the coherence is calculated using hashed vector values as input. Finally, the calculated coherence score measures the semantic relationship between words and will help find actionable insights. Coherence frequency is an important metric for extracting actionable insights from text data. The measure identifies key topics and semantic relationships, providing clear insights for business decision-making. The HV demonstrates greater resilience than alternative vectorizers, effectively managing empty documents, Out-of-Vocabulary (OOV) words, and hash collisions.

3) **Perplexity (PPl)**

Perplexity (PPL) is a popular metric for assessing the performance of a Language Model (LM), and is defined as the inverse of the probability of the test data being normalized by the number of words count [33]. PPL is the inverse geometric mean of subsequent word possibilities p(wiw<i) in a text with N signals(w1w2…wN) [34]. Perplexity can be calculated by using the following equation (3)

$$PPL = 2\hat{}(-\sum(p(x) * log2(p(x)))) \qquad (3)$$

where PPL is the perplexity, p(x) is the Probability distribution of the data, $\log_2$ is the base 2 logarithm and $\sum$ is the Summation of all data points

IV. THE PROPOSED METHODOLOGY

This research proposes a comprehensive framework for extracting actionable insights from the given dataset. This framed methodology efficiently extracts valuable insights for subsequent processes such as decision-making, opinion analysis, sentiment analysis, or recommending solutions for a procedure. A novel method known as the Hashed-Coherence Frequency Calculator (HCFC) has been developed to extract actionable insights from any provided dataset. The overall approach is picturized in Fig. 1.

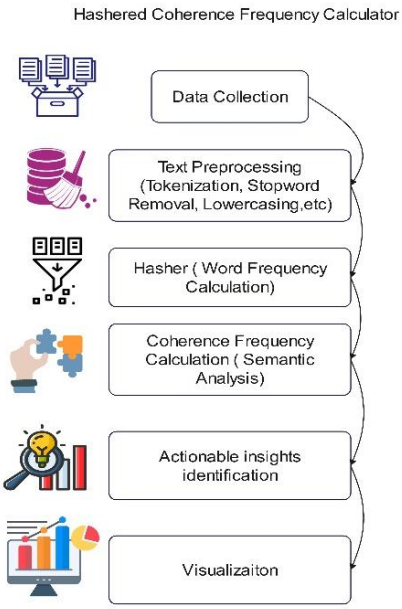


Fig. 1. Overall Methodology

The methodology shown in Fig. 1 contains 6 stages, and the first stage involves reading the data from the dataset and applying preprocessing over the read data. Subsequently, the preprocessed data is input into the Hasher, where word frequency calculation is performed, followed by the calculation of coherence frequency. Ultimately, actionable insights characterized by strong mutual bonding are identified and visualized. The picturized steps in Fig. 1 are implemented in the Hashed Coherence Frequency Calculator(HCFC) algorithm.

Algorithm Hashed Coherence Frequency Calculator (HCFC)

HCFC(D)

Step 1: Preprocess the Dataset D

a. Let D be the dataset and x be the text $\ni x \in D$

b. Preprocess the x

c. f(x) = tokenize(lowercase(remove_stopwords(x))))

where tokenize breaks x into words, lowercase converts x into lower case, and remove_stopwords removes all stop words from x.

Step 2: Calculate the word frequencies

a. Let V be the vocabulary, and h be the hashing function

$\ni$ h: V $\rightarrow$ R

where :

- V is the vocabulary (set of unique words)
- R is the set of real numbers
- h is the hashing function that maps each word in V to a unique real number

b. Calculate the Hashing matrix for each column. c

Hc = [h(w_1), h(w_2), ..., h(w_n)]

where:

- Hc is the hashing matrix for column c
- w_i is the i-th word in the vocabulary
- n is the number of words in the vocabulary
- h(w_i) is the hashed value of the i-th word

Step 3: Calculate the Coherence Frequency

a. Calculate the Coherence Frequency

$Cc = \sum(Hc)$

where

- $\sum$ denotes the sum of all words in the vocabulary
- Cc is the coherence frequency for column c

Step 4: Find the Actionable insights

a. Let T = 0.1 (denoting the threshold value)

b. Find actionable insight column A

$A = \{c \mid Cc > T\}$

where:

- A is the set of columns with actionable insights
- Cc > T indicates the coherence frequency of column c, which is above the threshold value.

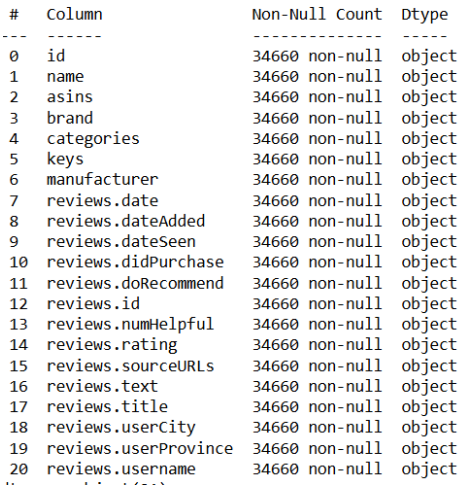
The Hashed Coherence Frequency Calculator (HCFC) algorithm uses a text-based analysis approach to compute the coherence frequency of words in the given dataset. The algorithm begins by preprocessing the text, transforming it to lowercase, eliminating stop words, and segmenting it into individual tokens. Subsequently, it computes the hashing matrix for each term, reflecting the term's frequency within the text. The approach then computes the coherence frequency for each word by aggregating the values of the hashing matrix. Ultimately, it discerns actionable insights by picking terms with coherence frequencies exceeding a specified threshold. The HCFC algorithm offers a quick and effective method for analyzing text data and identifying significant words and phrases to detect actionable insights.

V. RESULTS AND ANALYSIS

This section comprehensively analyses the outcomes of the Hashed Coherence Frequency Calculator (HCFC) algorithm for detecting actionable insight columns in the given datasets. The overall analysis reveals underlying patterns and trends in customer feedback, providing valuable insights for improving product design, app development, and customer satisfaction. The performance analysis is conducted through the following two case studies using the Amazon product and Google Play Store Apps review datasets.

A. Case Study-1

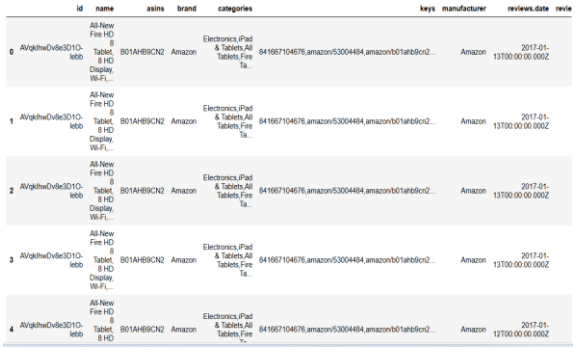
Case Study 1 uses a dataset from the Kaggle database[31] of about 34,000 consumer review details for Amazon items, such as Kindle and Fire TV Stick, acquired from Datafiniti's Product Database. The dataset contains 21 columns and is picturized in Fig .2.



#	Column	Non-Null Count	Dtype
0	id	34660 non-null	object
1	name	34660 non-null	object
2	asins	34660 non-null	object
3	brand	34660 non-null	object
4	categories	34660 non-null	object
5	keys	34660 non-null	object
6	manufacturer	34660 non-null	object
7	reviews.date	34660 non-null	object
8	reviews.dateAdded	34660 non-null	object
9	reviews.dateSeen	34660 non-null	object
10	reviews.didPurchase	34660 non-null	object
11	reviews.doRecommend	34660 non-null	object
12	reviews.id	34660 non-null	object
13	reviews.numHelpful	34660 non-null	object
14	reviews.rating	34660 non-null	object
15	reviews.sourceURLs	34660 non-null	object
16	reviews.text	34660 non-null	object
17	reviews.title	34660 non-null	object
18	reviews.userCity	34660 non-null	object
19	reviews.userProvince	34660 non-null	object
20	reviews.username	34660 non-null	object

Fig. 2. Columns in Amazon Consumer's Products review dataset

The columns in the Amazon customer product reviews depicted in Fig. 2 illustrate diverse aspects of the items and the shopping experience. Each row in the dataset denotes distinct feedback of an individual consumer. A sample from the consumer product review dataset is presented in Fig. 3 below.



	id	name	asins	brand	categories	keys	manufacturer	reviews.date	revie
0	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	
1	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	
2	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	
3	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	
4	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-12T00:00:00.000Z	

Fig. 3. Sample data from Amazon Consumer's Products review dataset

Applying the Hashed Coherence Frequency Calculator (HCFC) technique to the dataset in Fig.3 reveals the feedback data's underlying patterns and trends, discovering useful insights that can guide future product sales and decision-making. After applying the HCFC algorithm, the separated actionable insights from the dataset with 5000 samples are shown in Table. I, along with their coherence frequency count.

TABLE I SEGREGATION OF ACTIONABLE INSIGHTS WITH THEIR COHERENCE FREQUENCY

Top Actionable Insights columns		
S.No.	Column Name	Coherence frequency count
1.	reviews.sourceURLs	4593
2.	categories	4053
3.	name	1725
4.	reviews.title	1158
5.	reviews.text	512

Table I presents the columns featuring high coherent frequencies as actionable insights columns. The listed columns, reviews, sourceURLs, and categories are the most pertinent actionable insights, with

coherence frequencies of 4593 and 4053, respectively. Product names, review titles, and review text follow with subsequent significance.

TABLE II SEGREGATION OF OUTLIERS (NON-ACTIONABLE ITEMS) WITH THEIR COHERENCE FREQUENCY

Outliers(Non-Actionable items) Columns		
S.No.	Column Name	Coherence frequency count
1.	id	0
2.	asings	0
3.	brand	0
4.	keys	0
5.	manufacturer	0
6.	reviews.date	0
7.	reviews.dateAdded	0
8.	reviews.dateSeen	0
9.	reviews.didPurchase	0
10.	reviews.doRecommend	0
11.	reviews.id	0
12,	reviews.numHelpful	0
13.	reviews.Rating	0
14	reviews.userCity	0
15	reviews.userProvince	0
16	reviews.username	16

Table II lists segregated outliers (non-actionable items) and their coherence frequencies. After preprocessing, the HCFC algorithm counts the maximum number of pertinent words in a column as the coherence frequency. Before computing the coherence frequency, all textual data are transformed into numerical vectors via the hashed vectorizer function. The HCFC algorithm identifies actionable insights by focusing on columns with the highest coherence frequency, likely to contain relevant and meaningful information.

However, the lowest coherence frequency columns are less likely to contain valuable information and are regarded as non-actionable insights. Fig. 4 visualizes the plotting of the top actionable insights columns in the given dataset.

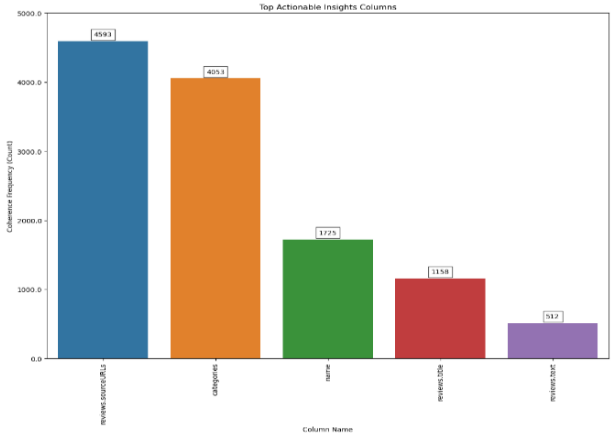


Fig.4 Top Actionable Insights from the Amazon Consumer Product Review dataset.

Each Actionable Insight column is represented as a separate bar with different colours and maximum coherence frequency at the top of each bar. The HCFC algorithm found a total of 5 actionable insights columns.

The picturization of non-actionable insights (outliers) is represented in Fig. 5.

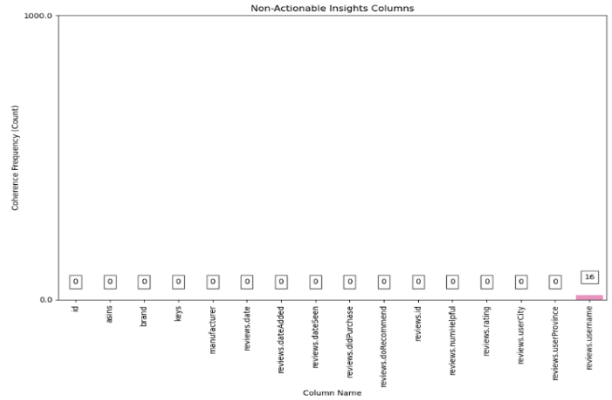


Fig.5 The outliers (non-actionable insights) in Amazon Consumer Product Review dataset.

Fig. 6 illustrates the complete classification of all columns, designated as either actionable insights or outliers (non-actionable insights), together with their highest coherence frequency bound.

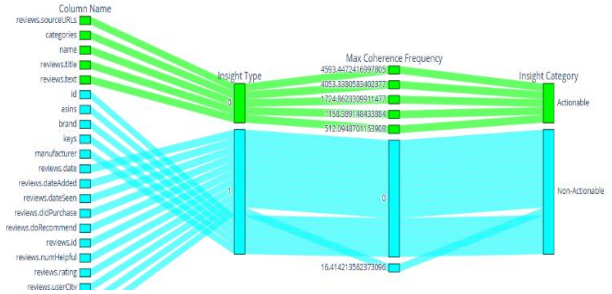


Fig.6 The complete grouping of actionable and non-actionable insights columns with their coherence frequency values

The perplexity plot for all actionable insight columns is plotted in Fig. 7

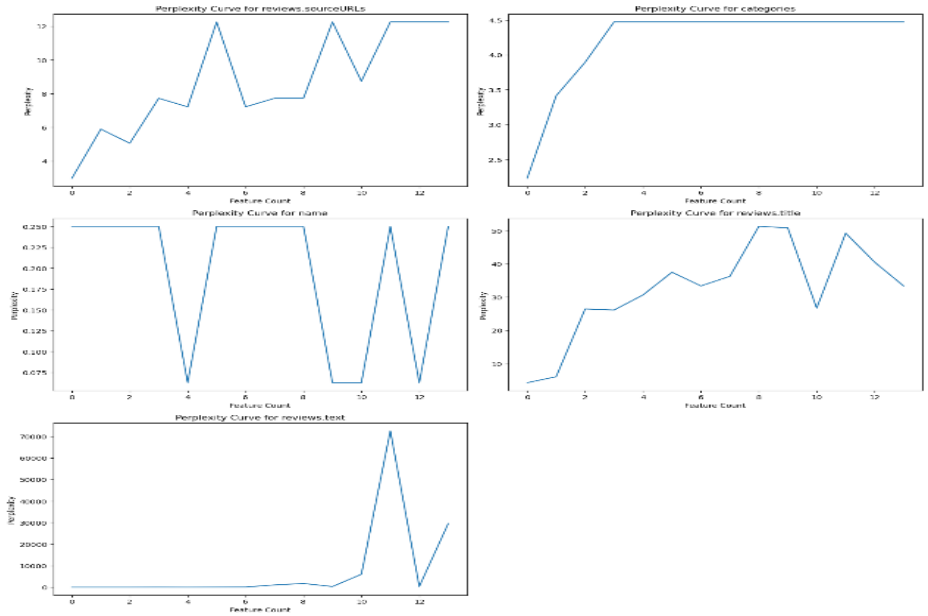
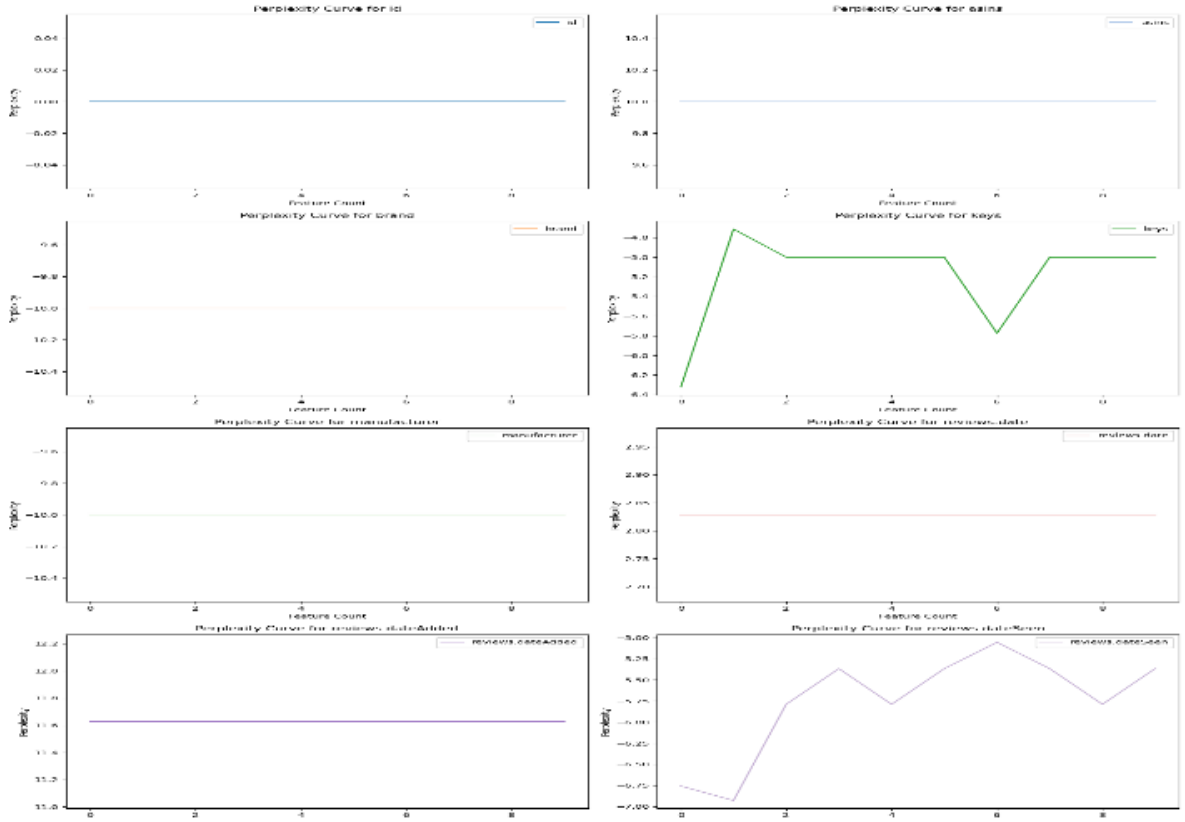


Fig.7 The perplexity plot for all actionable insight columns.

The produced curve shapes for each actionable insight column in Fig. 7 indicate its importance and binding to others. All the actionable insight columns have stable convergence at certain points, which may lead to good topic modeling detection. The decreasing level at certain points indicates the increasing feature count towards better topic modeling. From the observation, it is concluded that the HCFC predicted the exact actionable insights columns based on their importance.

The perplexity plot for all remaining non-actionable insight columns is presented in Fig. 8.



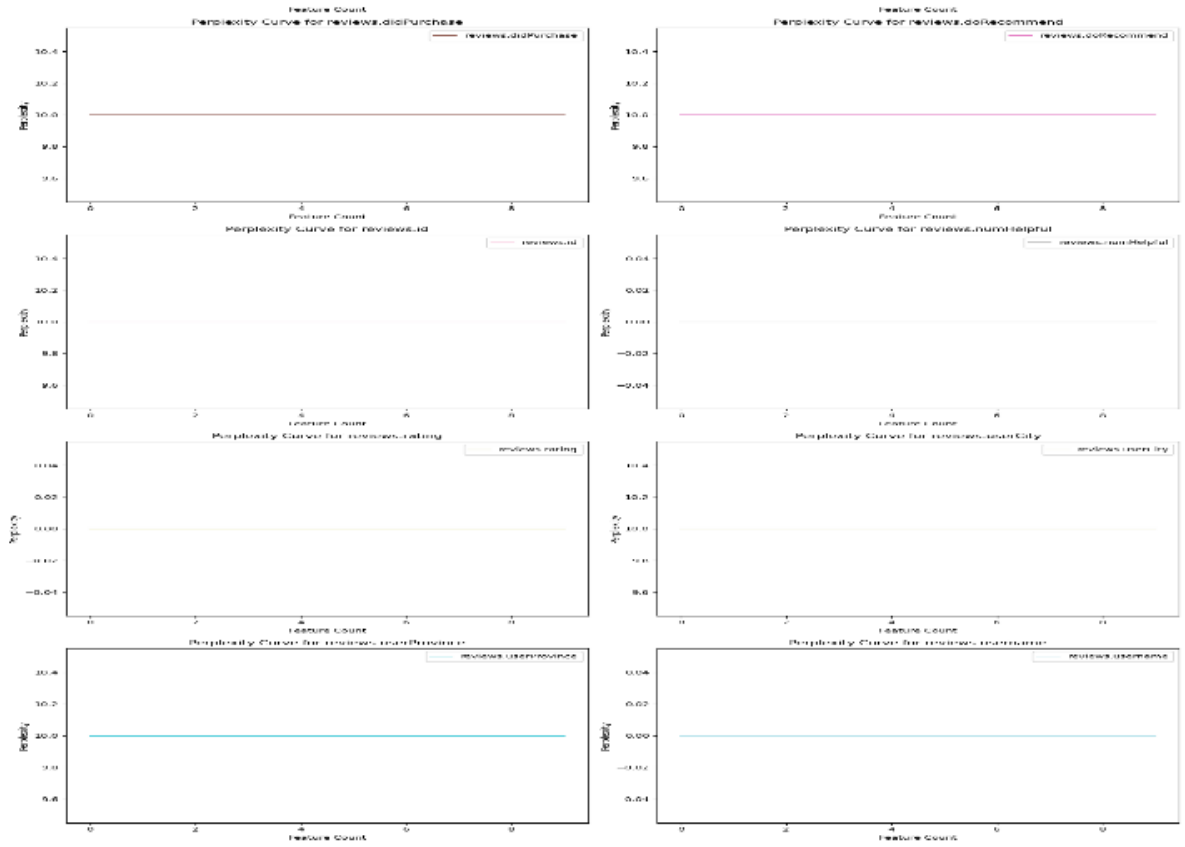


Fig. 8 The perplexity plot for all non-actionable insight columns.

The plotted perplexity curves for all non-actionable insights columns from Fig. 8 show the binding relation with other columns. Most columns have no stable convergence progression and flat curves, indicating no binding and no more feature count for topic detection.

B Case Study-2

Caste Study-2 used a dataset containing 12K review records of real users about the app at Google Play Store from the Kaggle database [32]. The dataset contains about 12 columns, which are depicted in Fig. 9.

```
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   reviewId              12495 non-null  object
 1   userName              12495 non-null  object
 2   userImage             12495 non-null  object
 3   content               12495 non-null  object
 4   score                 12495 non-null  object
 5   thumbsUpCount         12495 non-null  object
 6   reviewCreatedVersion  12495 non-null  object
 7   at                    12495 non-null  object
 8   replyContent          12495 non-null  object
 9   repliedAt             12495 non-null  object
 10  sortOrder             12495 non-null  object
 11  appId                 12495 non-null  object
```

Fig.9 Columns in the Play Store Reviews dataset

Tables III and IV present the actionable and non-actionable columns from the above dataset.

TABLE III SEGREGATION OF ACTIONABLE INSIGHTS WITH THEIR COHERENCE FREQUENCY

Top Actionable Insights columns		
S.No.	Column Name	Coherence frequency count
1.	sortOrder	5000
2.	content	789
3.	replyContent	284
4.	userName	49
5.	Reviewed	07

TABLE IV SEGREGATION OF OUTLIERS (NON-ACTIONABLE ITEMS) WITH THEIR COHERENCE FREQUENCY

Outliers(Non-Actionable items) Columns		
S.No.	Column Name	Coherence frequency count
1.	score	0
2.	thumbsUpCount	0
3.	reviewCreatedVersion	0
4.	at	0
5.	repliedAt	0
6.	appId	0
7.	userImage	1

Tables III and IV show that the HCFC algorithm dutifully identified the actionable and non-actionable insights from the Play Store dataset. In case study 2, the algorithm also used coherence frequency count over 5000 data samples to identify the actionable and non-actionable insight columns. The maximum occurrence of words in columns is counted as coherence frequency and subjected to insight identification. Fig. 10 plots the top actionable insights columns from the Play Store Review dataset as a bar chart with their coherence frequency count at the top of each bar.

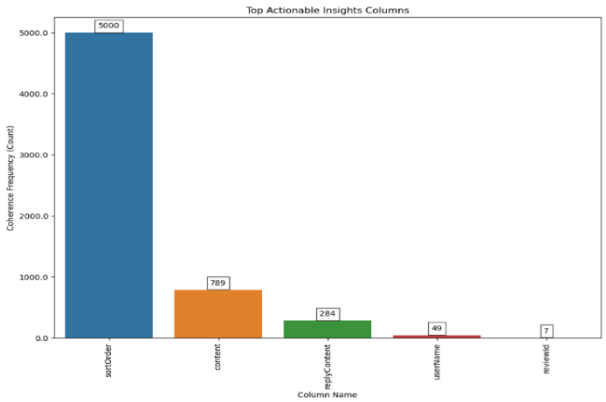


Fig 10 Top Actionable Insights columns from the Google Play Store Review Dataset

The bar plotting of non-actionable insight columns from the Play Store Review dataset is picturized in Fig. 11.

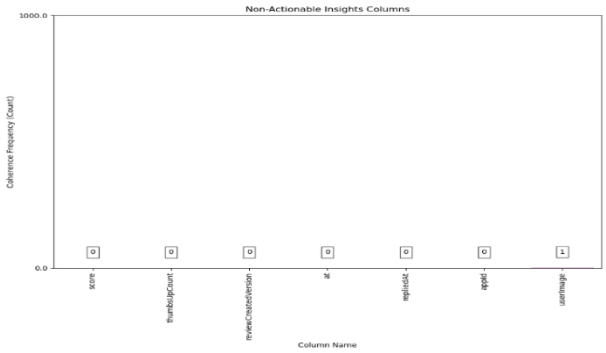


Fig 11 Non-Actionable insights columns in the Google Play Store Review Dataset

Fig. 12. shows the complete segregation of actionable and non-actionable insights columns and their maximum coherence frequency bound.

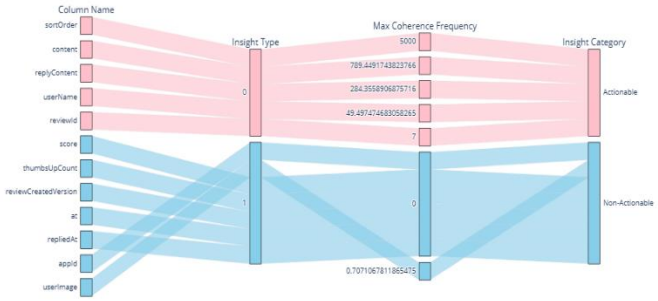


Fig 12 Complete visualization of actionable and non-actionable insight columns from Google Play Store Review Dataset.

Fig. 12 visualizes all actionable and non-actionable insights columns of the Play Store Review dataset with their coherence frequency values. The figure displayed the column names and their coherence frequencies as two groups in a clear visual presentation.

The perplexity plot for both Actionable and non-actionable insights columns is shown in Fig. 13 and Fig. 14.

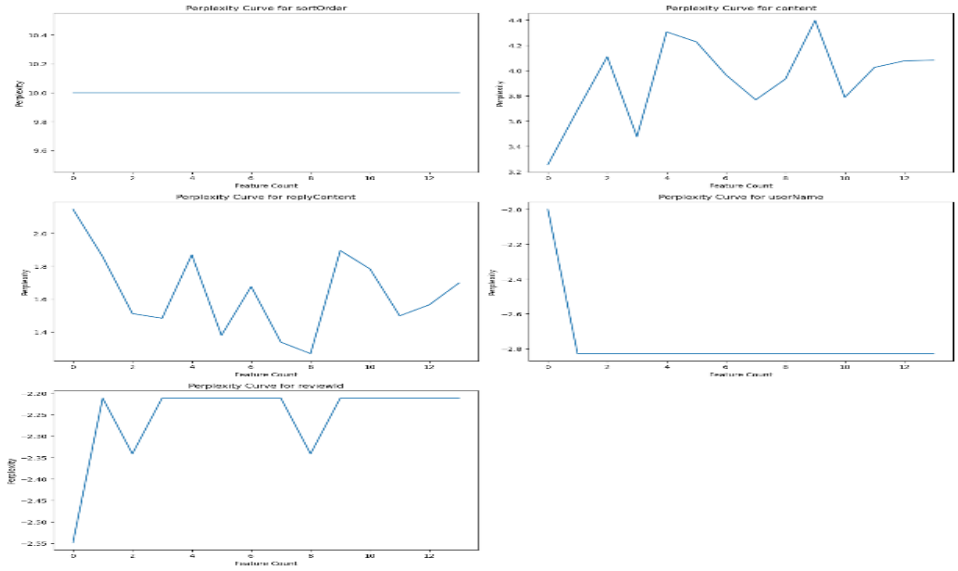


Fig 13 Perplexity plot for all actionable insight columns in Play Store Review Dataset

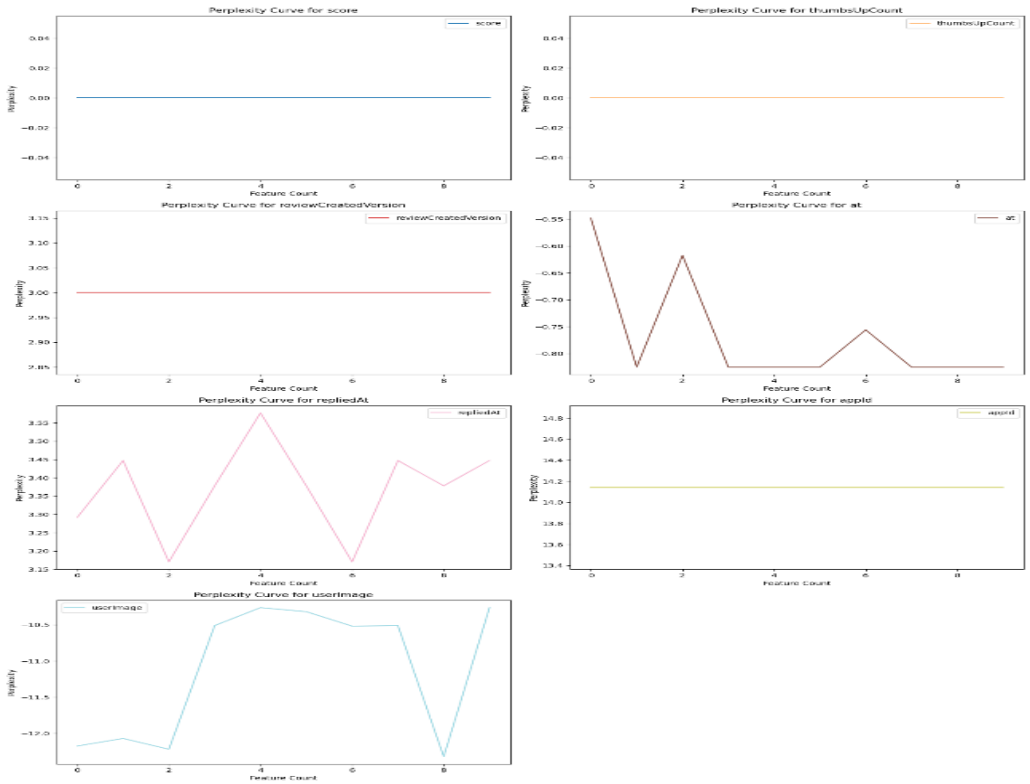


Fig 14 Perplexity plot for all non-actionable insight columns in Play Store Review Dataset

The perplexity plot for actionable insights columns in Fig. 13 shows that the curves converge for certain columns and are flat for one column. This indicates that the actionable insights columns have been assessed using the coherence frequency count, and the curve shape indicates their relationship with others. In Fig. 14, maximum columns have high peak values consecutively and flat shapes, indicating minimum binding with other columns tending to a topic model and with minimum coherence frequency counts.

C. Result Analysis and Discussion

The results obtained from both case studies 1 and 2 show that the proposed HCFC algorithm efficiently identified the actionable and non-actionable insights columns in the given dataset. The primary processing involved in this algorithm is preprocessing and the calculation of word frequencies. So, the time complexity of HCFC is O(n*m), where n is the number of texts and m is the average text length. The space complexity of the algorithm is also O (n * m), as it takes space for storing the dataset and vocabulary. The scalability of the algorithm is also having a notable benefit, allowing it to handle large datasets efficiently and with ease. As the number of samples grows, the hashing matrix will also grow linearly. This makes sure that the method stays as fast as it can be. This algorithm's scalability enables it to effectively tackle complex data analysis tasks, making it a strong candidate for big data applications. The performance analysis of the HCFC with TF-IDF, Word embedding, and LDA is illustrated in Fig. 15.

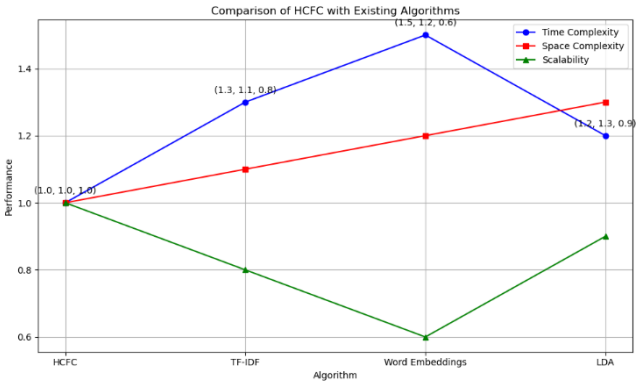


Fig 15 Performance Analysis of HCFC algorithm with TF-IDF, Word-Embedding, and LDA

Fig. 15 visualizes the performance analysis of the HCFC, TF-IDF, Word embedding, and LDA algorithms. Three measures, namely, Time Complexity, Space Complexity, and Scalability, are analyzed for all the algorithms considered, leading to the following conclusions.

- The HCFC algorithm has shown optimized performance and is appropriate for general topic detection, text classification, and real-time analysis applications. Its optimistic characteristics render it suitable for applications in edge computing, cloud computing, and real-time processing involving larger datasets.
- TF-IDF exhibits significant time complexity while demonstrating favorable space efficiency and scalability. Word embedding establishes high temporal efficiency while maintaining favorable spatial complexities. LDA discloses low time complexity, favorable space complexity, and optimal scalability.

The overall comparative analysis shows that the designed HCFC algorithm has balanced performance over all the used metrics, efficient algorithmic design, and low computational overhead. The algorithm can be applied to real-time processing like chatbot development, text classification, sentimental analysis, and topic modeling.

VI. CONCLUSION

The present research work developed an innovative algorithm known as the Hashed Coherence Frequency Calculator (HCFC), which serves as a comprehensive framework for distinguishing between actionable and non-actionable items within the provided data sources. Following the meticulous design and execution of the HCFC alongside two case studies, the findings demonstrated the effectiveness of the proposed algorithm in distinguishing between actionable and non-actionable insights. By utilizing coherence frequency count and hashing vectorizer, HCFC shows impressive performance, and its linear scalability guarantees computational efficiency. The comprehensive result analysis indicated that the developed algorithm is an optimal solution for applications involving vast quantities of data. The work can be improved by incorporating essential dimensionality reduction methods to enhance computational efficiency.

CONFLICT OF INTEREST: Only one author is involved in this work, so there is no conflict of interest throughout.

REFERENCE

[1] Sarker, Iqbal H. "Data science and analytics: an overview from data-driven smart computing, decision-making and

applications perspective." SN Computer Science 2, no. 5 (2021): 377.
[2] Kruspe, Anna, Jens Kersten, and Friederike Klan. "Detection of actionable tweets in crisis events." Natural Hazards and Earth System Sciences 21, no. 6 (2021): 1825-1845.
[3] Khan, Roman, Muhammad Usman, and Muhammad Moinuddin. "From raw data to actionable insights: navigating the world of data analytics." International Journal of Advanced Engineering Technologies and Innovations 1, no. 4 (2024): 142-166.
[4] Nivedhaa, N. "From Raw Data to Actionable Insights: A Holistic Survey of Data Science Processes." International Journal of Data Science (IJDS) 1, no. 1 (2024): 1-16.
[5] McCreadie, Richard, Cody Buntain, and Ian Soboroff. "Incident streams 2019: Actionable insights and how to find them." (2020): 744-760.
[6] Jansen, Bernard J., Joni O. Salminen, and Soon-gyo Jung. "Data-driven personas for enhanced user understanding: Combining empathy with rationality for better insights to analytics." Data and Information Management 4, no. 1 (2020): 1-17.
[7] Bogatu, Alex, Alvaro AA Fernandes, Norman W. Paton, and Nikolaos Konstantinou. "Dataset discovery in data lakes." In 2020 ieee 36th international conference on data engineering (icde), pp. 709-720. IEEE, 2020.
[8] Kumar, S., Sharma, A., and Singh, H. "Machine Learning for Business Analytics: A Review." Journal of Business Analytics 3, no. 1 (2020): 1-12.
[9] Narne, Harish. "AI-DRIVEN DATA ANALYTICS TRANSFORMING BIG DATA INTO ACTIONABLE INSIGHTS." INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING (IJAIML) 2, no. 01 (2023): 142-154.
[10] Shollo, Arisa, and Robert D. Galliers. "Constructing actionable insights: the missing link between data, artificial intelligence, and organizational decision-making." In Research Handbook on Artificial Intelligence and Decision Making in Organizations, pp. 195-213. Edward Elgar Publishing, 2024.
[11] Coussement, Kristof, and Dries F. Benoit. "Interpretable data science for decision making." Decision Support Systems 150 (2021): 113664.
[12] Luz, Ayuns, and Godwin Olaoye. Leveraging Data for Actionable Insights and Engagement Strategies. No. 13218. EasyChair, 2024.
[13] Gabelaia, Ioseb. "The Use of Artificial Intelligence to Convert Social Media Data into Actionable Insights." In International Conference on Reliability and Statistics in Transportation and Communication, pp. 167-178. Cham: Springer Nature Switzerland, 2023.
[14] Patel, Priya, and Sumit Gahletia. "Deriving Actionable Insights from Big Data to Enhance Customer Experiences Across the Consumer Journey." International Journal of Responsible Artificial Intelligence 10, no. 8 (2020): 1-9.
[15] Rao, A. Suresh, B. Vishnu Vardhan, and Hafeezuddin Shaik. "Role of exploratory data analysis in data science." In 2021 6th International Conference on Communication and Electronics Systems (ICCES), pp. 1457-1461. IEEE, 2021.
[16] Hullman, Jessica, and Andrew Gelman. "Designing for interactive exploratory data analysis requires theories of graphical inference." Harvard Data Science Review 3, no. 3 (2021): 10-1162.
[17] Manju, M. K., Abin Oommen Philip, and M. U. Sreeja. "Exploratory data analysis of Indian Premier League (IPL)." In AIP Conference Proceedings, vol. 2773, no. 1. AIP Publishing, 2023.
[18] Badmus, Oluwaseun, Shahab Anas Rajput, John Babatope Arogundade, and Mosope Williams. "AI-driven business analytics and decision making." (2024).
[19] Rajan, Preethi. "Integrating IoT analytics into marketing decision making: A smart data-driven approach." International Journal of Data Informatics and Intelligent Computing 3, no. 1 (2024): 12-22.
[20] Xu, Yuanzhi. "Customer Feedback Segmentation, Summarization, and Natural Language Querying Using Machine Learning." (2024).
[21] Al Mesmari, Saleimah. "Transforming data into actionable insights with cognitive computing and AI." Journal of Software Engineering and Applications 16, no. 6 (2023): 211-222.
[22] Zeng, E. Zhixuan, Hayden Gunraj, Sheldon Fernandez, and Alexander Wong. "Explaining Explainability: Towards Deeper Actionable Insights into Deep Learning through Second-order Explainability." arXiv preprint arXiv:2306.08780 (2023).
[23] Kharakhash, O. "Data visualization: transforming complex data into actionable insights." Automation of technological and business processes 15, no. 2 (2023): 4-12.

[24] Olaniyi, Oluwaseun, Nishant Hemantkumar Shah, Anthony Abalaka, and Folashade Gloria Olaniyi. "Harnessing predictive analytics for strategic foresight: a comprehensive review of techniques and applications in transforming raw data to actionable insights." Available at SSRN 4635189 (2023).
[25] Hasan, Emrul, Mizanur Rahman, Chen Ding, Jimmy Xiangji Huang, and Shaina Raza. "based Recommender Systems: A Survey of Approaches, Challenges and Future Perspectives." arXiv preprint arXiv:2405.05562 (2024).
[26] Yang, Yi, and Ramanath Subramanyam. "Extracting Actionable Insights from Text Data: A Stable Topic Model Approach." MIS Quarterly 47, no. 3 (2023).
[27] Tijare, Poonam, and P. Jhansi Rani. "Exploring popular topic models." In Journal of Physics: Conference Series, vol. 1706, no. 1, p. 012171. IOP Publishing, 2020.
[28] Abdelrazek, Aly, Yomna Eid, Eman Gawish, Walaa Medhat, and Ahmed Hassan. "Topic modeling algorithms and applications: A survey." Information Systems 112 (2023): 102131.
[29] Papadia, Gabriele, Massimo Pacella, Massimiliano Perrone, and Vincenzo Giliberti. "A comparison of different topic modeling methods through a real case study of italian customer care." Algorithms 16, no. 2 (2023): 94.
[30] Roshan, R., Bhacho, I.A. and Zai, S., 2023. Comparative Analysis of TF–IDF and Hashing Vectorizer for Fake News Detection in Sindhi: A Machine Learning and Deep Learning Approach. Engineering Proceedings, 46(1), p.5.
[31] https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products
[32] https://www.kaggle.com/datasets/prakharrathi25/google-play-store-reviews
[33] Lee, Nayeon, Yejin Bang, Andrea Madotto, and Pascale Fung. "Misinformation has high perplexity." arXiv preprint arXiv:2006.04666 (2020).
[34] Kuribayashi, Tatsuki, Yohei Oseki, Takumi Ito, Ryo Yoshida, Masayuki Asahara, and Kentaro Inui. "Lower perplexity is not always human-like." arXiv preprint arXiv:2106.01229 (2021).

ABOUT AUTHOR

Dr. B. Srinivasan is working as an Associate Professor of Computer Science at the School of Arts and Science, Vinayaka Mission's Chennai Campus, Paiyanoor , Vinayaka Mission's Research Foundation Deemed to be University, Salem, and has 26 years of academic and 17 years of research experience. His specialization areas are Artificial Intelligence, Neural Networks, Big Data Analytics, Machine Learning, Internet of Things (IoT) and Cloud Computing. He published research papers on recent topics and also wrote books on advanced topics.