Due Mar 23   45 points

13 Replies (2)   🔖 ⋮

Unit 1: Discussion

📢 Directions

Find an example of data mining in the news (e.g., it occurred in the last 3 months).

- Share a link to the article.
- Summarize the article.
- Discuss additional data mining approaches you might have taken.
- Debate what risks the analyst might have encountered.

Note: Once you have posted your initial response, please respond to the postings of **two of your classmates'** initial postings. Subsequent postings can be in whatever format you prefer. To earn full credit, be sure to ask questions or additional information that furthers discussion. Comments that are the equivalent of, "Great response, that's what I think" will not earn full credit.

📅 Criteria for Success

Initial Post (DUE: Thursday 11:59 p.m. CT)

- In the initial post you will do the following:
  - Uses the weekly materials to construct an academic argument that addresses the discussion question in a thorough and logical manner.

- Correctly uses key terms and concepts. Thoroughly addresses all components of the prompt. Ideas are clear and on-topic.
- Follows grammar conventions. The writing is concise and easy to read.
- Writes approximately 200 words.

Response to Two Peers (DUE: Sunday, 11:59 CT)

- In each response, you will do the following:
  - Respond to at least two student's discussion posts. In each response, recommend some additional data mining techniques that could have been used and address what risks the analysis might have faced.
  - Follows grammar conventions. The writing is concise and easy to read.
  - Writes approximately 100 words.

Please review the rubric for this assignment before beginning to ensure that you earn full credit. Contact me if you have any questions.

Reply

Avinash Bunga (https://canvas.park.edu/courses/85581/users/111811)
Mar 20 4:39pm

Avinash Bunga

Information Systems and Business Analytics, Park University

CIS625HOS2P2025 Machine Learning for Business

Professor: Abdelmonaem Jornaz

March 20, 2025

Unit 1: Discussion

Data Mining in Autonomous Driving

Summary of the Article: This article explains how data mining plays a big role in how self driving car technology is growing. It focuses on how autonomous vehicles (AVs) gather huge amounts of data through sensors like cameras, LiDAR, GPS, and radar and how they use this data to get smarter over time.

Link to the Article: **https://arxiv.org/abs/2401.12888** (https://arxiv.org/abs/2401.12888)

The authors describe three major stages of how data mining has evolved in this space:

1. **First Generation (2012–2016):** Simple image datasets (like KITTI and Cityscapes) helped vehicles recognize objects and understand the road around them.
2. **Second Generation (2019–2021):** Datasets started to include multiple sensors, like LiDAR and HD Maps, with platforms such as Waymo and NuScenes leading the way.
3. **Third Generation (2023–Now):** We now see advanced tools like Vision Language Models (VLMs) and Large Language Models (LLMs), helping AVs make better choices in tough or unusual driving situations (Kerner, n.d.; NVIDIA, n.d.; Zhang et al., 2024).

The article also highlights how companies like Tesla and NVIDIA use closed loop systems, where real world driving data is constantly collected and then used to further train and improve their models.

Additional Data Mining Approaches: Here are some other ways data mining could be used to make self-driving cars even better:

Approach	Description
Anomaly Detection	Use tools that find strange or rare driving situations the car has not seen before so it can react better.
Federated Learning	Let each car learn on its own and share its findings instead of sending all data to one place. This helps keep private data safe.
Synthetic Data	Create fake driving scenarios (like bad weather or animals on the road) using simulation tools, to teach the car how to handle rare events.
Bias Detection	Look for patterns in the data that might be unfair (like too much city driving), and fix them so the car works well in all places and conditions.

Potential Risks Faced by Analysts: The article also discusses the risks of analyzing all this data. If these risks are not handled well, the models may not be accurate or fair, and could even be unsafe. Analysts need to be careful with how they collect, label, and use data.

Risk	Description
Privacy Issues	Data like GPS locations or driver habits can be sensitive, and collecting it might break privacy laws.

Biased Data	If most of the data comes from city driving in good weather, the model might not work well in other conditions.
Lack of Rare Event Data	Some events (like a deer suddenly crossing the road) are rare, so there is not enough data to teach the car how to handle them.
Labeling Mistakes	With so much data, it is easy to make mistakes when labeling it, these mistakes can lead the car to learn the wrong thing.
Over Reliance on Simulations	If the car is trained mostly with fake/simulated data, it may not perform well when facing real situations it has not seen before.

(Zhang et al., 2024).

Conclusion

This article gives a clear picture of how data mining helps self driving cars get better and safer. It also shows why it is important to be careful with privacy, fairness, and the accuracy of our data. As this technology keeps growing, data mining will continue to play an important role in making AVs more reliable and trustworthy.

References

Kerner, S. M. (n.d.). *What are large language models (LLMs)?* TechTarget. Retrieved March 20, 2025, from https://www.techtarget.com/whatis/definition/large-language-model-LLM

NVIDIA. (n.d.). *Vision language models*. NVIDIA Glossary. Retrieved March 20, 2025, from https://www.nvidia.com/en-us/glossary/vision-language-models/

Zhang, W., Ma, Y., Wu, Z., Yu, Y., Huo, J., Zheng, C., ... & Qi, J. (2024, January 23). *Data-centric evolution in autonomous driving: A comprehensive survey of big data system, data mining, and closed-loop technologies*. arXiv. https://arxiv.org/abs/2401.12888

Reply | ✉

Licurgo Silveira Teixeira (https://canvas.park.edu/courses/85581/users/119244)
Mar 20 3:26pm

[**Raymond James maintains strong buy on Digital Realty stock By Investing.com** ⬀](https://www.investing.com/news/analyst-ratings/raymond-james-maintains-strong-buy-on-digital-realty-stock-93CH-3906814)
[(https://www.investing.com/news/analyst-ratings/raymond-james-maintains-strong-buy-on-](https://www.investing.com/news/analyst-ratings/raymond-james-maintains-strong-buy-on-)
digital-realty-stock-93CH-3906814)

Published on March 20, 2025, the Investing.com article reports that Raymond James, a financial analysis firm, has maintained its "Outperform" recommendation and a $190 price target for Digital Realty Trust, Inc., a leading data center company. This sustained price target reflects Raymond James' confidence in the company's future performance, driven by the rising demand for data center infrastructure, particularly fueled by the artificial intelligence (AI) boom. While the article does not specify the methods used by Raymond James, stock analyses like this often rely on data mining to evaluate market trends, financial performance, and macroeconomic factors.

Raymond James could explore other data mining techniques to refine the evaluation:

1. Social Media Sentiment Analysis: Mine posts on social medias and platforms to gauge public perception of Digital Realty and the data center sector, especially regarding AI. This could leverage natural language processing (NLP) to identify positive or negative sentiments.
2. Time Series Forecasting: Analyze historical stock price data and trading volumes for Digital Realty, combined with tech market indicators (e.g., Nasdaq), to predict future movements with greater accuracy.
3. Competitive Data Mining: Compare Digital Realty's financial and operational metrics with competitors (e.g., Equinix) using clustering or regression analysis to pinpoint competitive advantages or hidden risks.
4. Correlation with AI Demand: Integrate data on energy consumption and data center capacity with trends in AI adoption by tech companies, using association algorithms to quantify the impact on Digital Realty's growth.

Risks the Analyst Might Have Encountered:

1. Incomplete or Biased Data: If Raymond James relied solely on Digital Realty's financial reports without cross-checking alternative sources, the data might overlook risks like hidden debt or rising operational costs, leading to an overly optimistic assessment.
2. Market Volatility and false correlation: The tech sector, particularly tied to AI, is prone to rapid shifts. A data mining model based on historical data might fail to predict sudden drops, such as those triggered by the tariff hikes from the Trump-era external trade policy model, reintroduced or speculated upon in early 2025. These policies could increase costs for tech infrastructure, impacting data center profitability and stock performance, especially for companies like Digital Realty that rely on global supply chains and energy-intensive operations.
3. Reputation and External Pressure: If the $190 forecast fails to materialize, Raymond James' credibility could be challenged, particularly if investors incur losses. Notably, Investing.com's own widely-used technical analysis sets a target price of $186.38, slightly below $190, and despite its buy recommendation, the site's market sentiment

and technical analysis suggest selling. A monthly chart analysis over the past five years shows the $190 price was hit in September 2024, indicating it's a significant level. Furthermore, in a post-pandemic recovery from 2020 to 2021, REITs (real estate investment trusts) like Digital Realty often rose when the broader stock market declined. However, during 2022 and 2023, as stocks regained strength, Digital Realty's price fell, reflecting an inverse trend. In the current economic policy shift, where tech-heavy markets like Nasdaq have dragged valuations down, tech-focused REITs are also declining alongside them, contradicting historical patterns. This inconsistency is driven by uncertainty and market sentiment, compounded by a short-term capital flight to European, Asian, and even smaller markets like Brazil as investors seek protection amid this volatile period. That said, Digital Realty's current price sits in a range previously tested multiple times as resistance or a supply zone, which could now act as a demand zone, offering potential support.

s

Reply | ✉

KK **Kouame Hermann Kouame**
(https://canvas.park.edu/courses/85581/users/123444)
Mar 20 3:04pm

1)Hello Class, as part of the discussion for Week 1 of our class, here is the discussion post I've created summarizing my article. The article is about the role of transcriptomic data, which is a type of data mining, in advancing cancer research, particularly through high-throughput sequencing technologies like RNA sequencing (RNA-seq) and here is the link: https://pmc.ncbi.nlm.nih.gov/articles/PMC11660120/

2)Cancer research has been revolutionized by the integration of transcriptomic data through high-throughput sequencing technologies like RNA sequencing (RNA-seq). This paper reviews the transformative impact of transcriptomics on understanding cancer biology, emphasizing the use of extensive datasets such as The Cancer Genome Atlas (TCGA) and

Genotype-Tissue Expression called (GTEx). While transcriptomic data provides crucial insights into gene expression patterns and disease mechanisms, the analysis is fraught with technical and biological biases. Technical biases include issues related to microarray, RNA-seq, and nanopore sequencing methods, while biological biases arise from factors like tumor heterogeneity and sample purity. Additionally, misinterpretations often occur when correlational data is erroneously assumed to imply causality or when bulk data is misattributed to specific cell types. This review emphasizes the need for researchers to understand and mitigate these biases to ensure accurate data interpretation and reliable clinical outcomes. By addressing these challenges, the paper aims to enhance the robustness of cancer research and improve the application of transcriptomic data in developing effective therapies and diagnostic tools. The integration of multi-omics approaches, such as proteomics and metabolomics, is also highlighted as essential for a more holistic understanding of cancer biology. The paper underscores the importance of recognizing and addressing biases in transcriptomic data mining to enhance the reliability and impact of cancer research, ultimately contributing to better cancer diagnosis, treatment, and prognosis.

2)To build upon transcriptomic data in cancer research, we can also utilize a range of supplementary data mining techniques such as integrating multi-omics data provides a more comprehensive view of cancer. Single-cell and spatial transcriptomics illuminate cellular heterogeneity and tumor architecture. Deconvolution methods can be improved to accurately identify cell types from bulk data. Network analysis and systems biology uncover key regulatory pathways. Machine learning and AI enable predictive modeling(Yang et al., 2021). Cross-disease studies reveal shared genetic mechanisms. Temporal studies track gene expression changes over time. Mendelian randomization and genetic association studies identify causal variants. These combined approaches address current limitations, aiming to enhance biomarker discovery, therapeutic targeting, and patient outcomes.

3) Analysts working with transcriptomic data in cancer research face several risks that could compromise their findings. Technical biases, such as PCR amplification errors in RNA-seq or high error rates in nanopore sequencing, can lead to inaccurate gene expression measurements. Biological biases, like tumor heterogeneity and sample contamination from non-cancerous cells, can obscure true tumor-specific signals. Misinterpreting correlations as causality may result in incorrect assumptions about gene functions, while misattributing bulk data to specific cell types can lead to flawed conclusions about gene roles in cancer(Sweeney et al., 2023). Additionally, discrepancies between mRNA levels and protein function due to factors like splice variants or post-translational modifications can misguide interpretations. Survivor bias in sample collection, where aggressive or metastatic cells are underrepresented, may also skew results. To mitigate these risks, analysts must employ robust normalization techniques, integrate multi-omics approaches, and leverage advanced methods like single-cell sequencing to ensure accurate and reliable conclusions.

Sources:

Sweeney, S. M., Hamadeh, H. K., Abrams, N., Adam, S. J., Brenner, S., Connors, D. E., Davis, G. J., Fiore, L., Gawel, S. H., Grossman, R. L., Hanlon, S. E.,   Hsu, K., Kelloff, G. J., Kirsch, I. R., Louv, B., McGraw, D., Meng, F., Milgram, D., Miller, R. S., . . . Srivastava, S. (2023). Challenges to using big data in   cancer. *Cancer Research*, *83*(8), 1175–1182. **https://doi.org/10.1158/0008-5472.can-22-1274** ⇱ **(https://doi.org/10.1158/0008-5472.can-22-1274)**

Yang, C., Yu, R., Ji, H., Jiang, H., Yang, W., & Jiang, F. (2021). Application of data mining in the provision of in-home medical care for patients with   advanced cancer. *Translational Cancer Research*, *10*(6), 3013–3019. https://doi.org/10.21037/tcr-21-896

Reply | ✉️

Battulga Bolormaa (https://canvas.park.edu/courses/85581/users/68062)
Mar 20 2:03pm

Hello class,

In a recent article by Reuters (2024), Wall Street investors have been using alternative data mining techniques to predict retail performance, particularly during the holiday shopping season. Traditionally, analysts rely on financial reports, but now they incorporate credit card transactions, consumer sentiment from surveys, and social media trends to forecast sales more accurately. Companies such as HundredX and Facteus provide this alternative data, allowing investors to gain early insights into consumer behavior. While some retail executives are concerned about the potential misinterpretation of data, retailers themselves are also using these data-driven methods to analyze their own business performance. This trend highlights the increasing role of data mining in financial decision-making and business strategy.

Additional Data Mining Approaches

While the article focuses on credit card transactions and social media analysis, other data mining techniques could enhance predictions. One method is web scraping, where analysts collect online reviews and pricing trends to gauge consumer interest. Additionally, machine learning algorithms can analyze historical sales data to identify patterns and predict future demand. Another approach is network analysis, which examines supply chain relationships to assess potential disruptions or advantages in product availability. Combining these methods with traditional financial metrics would provide a more comprehensive analysis of retail performance.

Potential Risks in Data Analysis

Despite its advantages, data mining presents several risks. First, there is the issue of data accuracy. If alternative data sources are unreliable or misinterpreted, they could lead to incorrect investment decisions. Second, privacy concerns arise when analyzing consumer transactions and online behavior, potentially violating ethical guidelines. Additionally, bias in the data could skew predictions. For example, social media sentiment analysis might favor

certain demographics, leading to misleading conclusions. Lastly, over-reliance on alternative data without considering macroeconomic factors like inflation or consumer confidence could result in flawed forecasts.

In conclusion, while data mining offers significant benefits in predicting retail trends, analysts must carefully assess data quality, ethical implications, and external factors to make well-informed decisions.

References

Reuters. (2024, November 25). Investors mining new data to predict retailers' results. Retrieved from https://www.reuters.com/business/retail-consumer/investors-mining-new-data-predict-retailers-results-2024-11-25/

Reply | ✉

Michael Oduro (https://canvas.park.edu/courses/85581/users/112167)   ⋮

Mar 20 1:26pm

Medusind Inc., a healthcare and medical billing company, experienced a significant data breach on December 29, 2023. This breach compromised the sensitive personal and health information of over 360,000 individuals. The stolen data included health insurance details, payment information, medical records, and government identification numbers.

On January 7, 2025, Medusind Inc. issued a public disclosure and sent notice letters and emails to the affected individuals. The company has since launched an investigation with cybersecurity experts and notified affected individuals.

Additional Data Mining Approaches:

Pattern Recognition – Analyzing network traffic to identify unusual patterns.

Machine Learning Models – Training models to detect anomalies in access logs.

Metadata Analysis – Examining file attributes and timestamps for signs of tampering.

Cross-Referencing External Breaches – Checking dark web databases for leaked data.

Potential Risks:

False Positives/Negatives – Misidentifying normal activity as a breach.

Data Privacy Violations – Improper handling of sensitive data during analysis.

System Overload – Intensive data mining may disrupt regular operations.

Legal Exposure – Missteps could lead to non-compliance with regulations (e.g., HIPAA).

Reference

Business Insider. (n.d.). *Medusind, Inc.. data breach under investigation by Levi & Korsinsky, LLP*. Business Insider. https://markets.businessinsider.com/news/stocks/medusind-inc-data-breach-under-investigation-by-levi-korsinsky-llp-1034237013?form=MG0AV3 ⤷ (https://markets.businessinsider.com/news/stocks/medusind-inc-data-breach-under-investigation-by-levi-korsinsky-llp-1034237013?form=MG0AV3)

Reply | ✉•

JM

Joseph Maina (https://canvas.park.edu/courses/85581/users/118606)

Mar 20 12:52am

⋮

Hello class,

Exploring Trends and Topics of Startup Research Over 25 Years Using Bibliometric and Topic Modeling Analysis

Source: https://prosperitas.uni-bge.hu/wp-content/prosperitas-upload/exploring-trends-and-topics-of-startup-research-over-25-years-using-bibliometric-and-topic-modelling-analysis-257.pdf ⤷ (https://prosperitas.uni-bge.hu/wp-content/prosperitas-upload/exploring-trends-and-topics-of-startup-research-over-25-years-using-bibliometric-and-topic-modelling-analysis-257.pdf)

Summary of the Article

This article analyzes the trends and topics of startup research over 25 years using bibliometrics and topic modeling analysis. When a startup proposes a business opportunity the aim is to produce and launch a groundbreaking service or product that will shake up the market. This analysis examines 1,195 papers from the Web of Science to identify publication trends, key contributing countries and journals, and the most researched startup-related topics. Topic modeling shows the top 10 major themes, including funding, innovation, success and failure, and the role of technology. This study highlights the increase of academic interest in startups that have a specific interest, particularly after the internet revolution, and provides insights for policymakers, researchers, and entrepreneurs regarding startup ecosystem development. Two unique methodologies are used in this study, quantitative analysis and topic modeling, to provide vital insights into startup research.

Additional Approaches through Data Mining

This study used bibliometric analysis and topic modeling as its main methodology of analysis, an alternative data mining approach could enhance the analysis:

1. **Sentiment Analysis**—Natural language processing (NLP) can be used to analyze sentiment trends in research papers, which could reveal and give insight into startups' perspectives on evolution.
2. **Clustering Techniques**—Unsupervised learning methods like k-means clustering can be used to group papers based on similarities of predefined topics to uncover underlying insights.
3. **Trend Prediction** – Applying time series analysis and predictive modeling to forecast future research directions in startup studies.

Challenges

1. **Data Bias** – The data being used for this study is English-language papers from the Web of Science database, it may overlook important research published in other languages or databases.
2. **Topic Modeling Limitations** – limitations of LDA in identifying topics and proposed a new algorithm-based topic modeling assumes fixed distributions of topics, which may oversimplify complex research themes.
3. **Temporal Interval** – The research trends may not necessarily reflect real-world changes in the startup ecosystem due to publication delays and dataset limitations.
4. **Overfitting** – If topic modeling parameters are not carefully tuned, the algorithm may generate topics that do not truly reflect the underlying research themes.

References

Baek, C., & Doleck, T. (2022). Educational Data Mining: a bibliometric analysis of an emerging field. *IEEE Access*, *10*, 31289–31296. **https://doi.org/10.1109/access.2022.3160457** ⤷ **(https://doi.org/10.1109/access.2022.3160457)**

Onan, A. (2019). Two-Stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. *IEEE Access*, *7*, 145614–145633. **https://doi.org/10.1109/access.2019.2945911** ⤷ **(https://doi.org/10.1109/access.2019.2945911)**

Reply | ✉

Jagadeesh Korukonda (https://canvas.park.edu/courses/85581/users/116942)
Mar 19 9:24pm

Hello Class,

Last week, Celonis, a German business "process mining" software company, sued SAP, Europe's largest software firm, in a federal court in San Francisco. The lawsuit accuses SAP of monopolising the business software market by limiting third-party vendors' access to customer data held within SAP applications, impacting billions of dollars of commerce. Celonis argues that SAP's actions make it harder and more expensive for them to obtain client data, thereby infringing on U.S. antitrust laws. The firm is requesting monetary compensation and an end to SAP's alleged anticompetitive activities.

Other Data Mining Techniques:

Apart from process mining, organizations can also apply various data mining techniques to enhance their processes:

- **Predictive Analytics**: Employing historical data to foretell future trends, aiding in decision-making as well as strategic planning.
- **Anomaly Detection:** Identification of unusual patterns or outliers in data, which may be valuable for fraud detection as well as quality control.
- **Association Rule Learning**: Discovering useful associations among variables in large databases, commonly used in market basket analysis to learn product purchase combinations.
- **Clustering:** Splitting up a collection of objects in such a way that objects in a group are more similar to each other than to members of another group, which is useful in customer segmentation.

Possible Hazards Encountered by Analysts:

Data miners must steer clear of several pitfalls:

- **Data Privacy Concerns:** Data privacy issues require strict compliance with privacy laws and regulations to prevent the possibility of unauthorized access or data breaches.
- **Data Quality Issues:** inaccurate, missing, or contradictory data can lead to misleading analysis and incorrect conclusions.
- **Ethical Considerations:** Ethical considerations need data mining processes to make sure they do not result in discrimination or biased results, which are essential for fairness and integrity.
- **Legal Risks:** Celonis's lawsuit underscores that strict data management policies can create legal uproar and accusations of anticompetitive practices.

By embracing robust data governance frameworks and being open about data practices, analysts are able to mitigate these risks and harness the full power of data mining techniques.

References:

Scarcella, M. (2025, March 14). *SAP sued by tech company Celonis over business process software*. Reuters. Retrieved March 19, 2025, from
https://www.reuters.com/legal/litigation/sap-sued-by-tech-company-celonis-over-

[business-process-software-2025-03-14/](https://www.reuters.com/legal/litigation/sap-sued-by-tech-company-celonis-over-business-process-software-2025-03-14/) ↪ (https://www.reuters.com/legal/litigation/sap-sued-by-tech-company-celonis-over-business-process-software-2025-03-14/)

Dutta, S. (2024, June 25). *Major issues of data mining: Navigating challenges and exploring solutions*. Sprinkle Data. https://www.sprinkledata.com/blogs/major-issues-of-data-mining-navigating-challenges-and-exploring-solutions ↪ (https://www.sprinkledata.com/blogs/major-issues-of-data-mining-navigating-challenges-and-exploring-solutions)

Reply | ✉

SS **Selorm Kwaku Soga** (https://canvas.park.edu/courses/85581/users/73415)
Mar 19 8:19pm | Last edited Mar 19 11:33pm

Hello Everyone,

Recent news of data mining comes from the ongoing debate around AI (artificial intelligence) systems using copyrighted creative materials without having sought permission.

Link To Article:

[https://www.thetimes.com/comment/the-times-view/article/artificial-intelligence-music-books-copyright-pppl0kmkp?utm_source=&region=global](https://www.thetimes.com/comment/the-times-view/article/artificial-intelligence-music-books-copyright-pppl0kmkp?utm_source=chatgpt.com&region=global) ↪ (https://www.thetimes.com/comment/the-times-view/article/artificial-intelligence-music-books-copyright-pppl0kmkp?utm_source=chatgpt.com&region=global)

Article Summary:

The "Make It Fair" campaign raises concerns that the UK government "is planning to allow AI companies to use the publicly available copyrighted content without permission under a 'text and data mining exemption'." This exemption will undermine the nation's £120 billion creative sector by weakening copyright protections and could damage economic and artistic output. Numerous artists and cultural figures have cautioned that such policies could quash creativity and diminish original works. Proposed alternatives include mandatory attribution for AI companies that scrape content from the Internet, to ensure transparency and fairness in the development of AI.

Additional Data Mining Approaches:

Data mining is a vital aspect of responsibly building AI in a copyright-compliant manner, as datasets must be legally acquired by establishing licensing contracts with the creators and publishers responsible for producing content to permit lawful access to datasets while still safeguarding intellectual property. Public domain and open-access resources (e.g., government publications) can also be used to obtain data without legal concerns (although restrictions may apply). Another ethical approach that minimizes these risks is the generation of synthetic data, where datasets generated artificially resemble real-world data.

Potential Risks Encountered by Analysts:

Data mining in AI analysts has many risks. Legally, if you use copyrighted content prospectively, you may face lawsuits, financial penalties, and injunctions. Reputational damage may also occur for organizations if content is perceived as being misused with past and current stakeholders losing trust. On an ethical level, scraping content without consent can put a significant strain on relationships with content creators. In addition, failure to comply with evolving data usage laws can lead to regulatory penalties. Analysts can also mitigate these risks by following ethical guidelines, obtaining proper licensing, assuring data quality, and providing transparency in their data collection and usage practices.

Reference:

Articles, T. T. L. (2025, February 24). *Ai must not be used to plunder copyrighted creative material*. The Times & The Sunday Times. **https://www.thetimes.com/comment/the-times-view/article/artificial-intelligence-music-books-copyright-pppl0kmkp?utm_source=&region=global** ▣ **(https://www.thetimes.com/comment/the-times-view/article/artificial-intelligence-music-books-copyright-pppl0kmkp?utm_source=chatgpt.com&region=global)**

Ibm. (2025, February 7). *What is data mining?*. IBM **https://www.ibm.com/think/topics/data-mining** ▣ **(https://www.ibm.com/think/topics/data-mining)**

Reply | ✉•

RN **Robert Nyabiti (https://canvas.park.edu/courses/85581/users/93498)**   ⋮
Mar 19 4:17pm | Last edited Mar 19 4:49pm

Article:
Designing Safe and Relevant Generative Chats for Math Learning in Intelligent Tutoring Systems. https://doi.org/10.5281/zenodo.14751365

Summary:
The article by Levonian et al. (2025) discusses the challenges faced when using Artificial Intelligence (AI) in math learning. The authors elaborate on how Large Language Models (LLMs) can present difficulties in tutoring despite their benefits, such as flexibility and availability. The data for the study was gathered from approximately 8,000 students who utilized Intelligent Tutoring Systems (ITSs). The authors designed a structured conversation system that employed open-ended questions and answers to evaluate both the system's performance and the students' responses. Their findings indicated that the structured system was effective in capturing inappropriate messages and responses from students. The authors recommended future research focused on "evaluating the effectiveness of LLMs in real ITSs" (Levonian et al., 2025, p. 85).

Data Mining Approaches:
The study by Levonian et al. (2025) could benefit by employing sentiment analysis as an

additional data mining tool. Sentiment analysis would help collect students' opinions about the AI, predict student trends, and better align with their needs (Wankhade et al., 2022). Additionally, the researchers could utilize text mining and natural language processing (NLP) techniques.

Risks:

The authors noted several limitations, including students' ability to game the system, the usage of only one LLM-backed design system within a single ITS, and challenges in following up on questions due to the single-response design approach.

References

Levonian, Z., Henkel, O., Li, C., & Postle, M.-E. (2025). Designing safe and relevant generative chats for math learning in intelligent tutoring systems. Journal of Educational Data Mining, 17(1), 66-97. **https://zenodo.org/records/14751365** ➦ **(https://zenodo.org/records/14751365)** .

Wankhade, M., Rao, A. C., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55, 5731-5780. **https://doi.org/10.1007/s10462-022-10144-1.** ➦ **(https://doi.org/10.1007/s10462-022-10144-1.)**

Reply | ✉

IK

Ian Koskei (https://canvas.park.edu/courses/85581/users/122159)
Mar 19 12:49pm | Last edited Mar 19 1:07pm

⋮

Unit 1 Discussion

A recent example of data mining in the news is Walmart's use of artificial intelligence (AI) to process massive volumes of product data to improve the experiences of associates and customers. In August 2024, Walmart announced its plan to utilize generative AI to analyze approximately 850 million product data points (Wassel, 2024). The goal of this initiative is to enhance retail experience through improved inventory control, personalized recommendations, and operational efficiency, hence leading to higher customer satisfaction and streamlined processes.

To further optimize its operations, Walmart could implement additional data mining techniques such as customer sentiment analysis and predictive analytics for inventory management. Zendesk (2024) states that customer sentiment analysis is when a company uses automation to examine the feedback customers leave in surveys, social media posts, and more. By examining customer reviews and feedback, sentiment analysis can provide insights into areas that need improvement and measure customer satisfaction. Predictive analytics, on the other hand, can help optimize stock levels and minimize overstock issues by forecasting demand based on past sales data.

However, implementing these data mining techniques comes with challenges, including data privacy concerns and data quality issues. Privacy compliance refers to adhering to laws and regulations that protect personal data (Karadeniz, 2024). Handling vast amounts

of customer data requires strict adherence to privacy regulations to protect personal information. Additionally, inaccurate or incomplete data can negatively impact decision-making processes, leading to incorrect conclusions.

References

Bryan Wassel. (2024). Walmart used AI to crunch 850M product data points. **https://www.retaildive.com/news/walmart-generative-ai-product-data-points/724782/** **(https://www.retaildive.com/news/walmart-generative-ai-product-data-points/724782/)**

Onur Karadeniz. (2024). Data Privacy and Compliance: Ensuring Safety in the Digital Age. **https://nmqdigital.com/blog/data-privacy-and-compliance-ensuring-safety-in-the-digital-age#:~:text=Privacy%20compliance%20refers%20to%20adhering,new%20and%20evolvin** **(https://nmqdigital.com/blog/data-privacy-and-compliance-ensuring-safety-in-the-digital-age#:~:text=Privacy%20compliance%20refers%20to%20adhering,new%20and%20evolving%20priva** .

Zendesk. (2024). Customer sentiment: What it is and why you need to measure it. **https://www.zendesk.com/blog/customer-sentiment/** **(https://www.zendesk.com/blog/customer-sentiment/)**

Reply | ✉

Ashish Thapa (https://canvas.park.edu/courses/85581/users/79401) ⋮
Mar 18 10:31pm │ Last reply Mar 20 2:03pm

Hello Everyone,

For this week's data mining discussion, I have chosen Levi's, one of the most famous jeans and apparel companies worldwide, and how they used techniques in data mining for their business benefit. In January 2025, Levi Strauss & Co. used data mining to figure out that baggy jeans were becoming popular again. They collected and analyzed huge amounts of data from online shopping records, online browsing records and searches, sales data from outlets and retail stores, and loyalty programs. Google's cloud machine learning capabilities were used for the data from 110 countries and 50,000 distribution points which in conclusion found a trend that more people wanted baggy jeans. After finding out what was in demand, Levi's adjusted their marketing strategies and made new designs, and stocked up on baggy designs to match what was the demand in the market and as a result, the sales of the loose-fitting denim went up by 15%, leading to profits. (Wsj, 2025) This week's materials we studied define data mining as" the exploration and analysis of the data to find patterns that enable us to make better decisions" which aligns with the approach Levi's made by analyzing consumer preferences rather than just predicting sales. Also as data mining is different from predictive modeling as it focuses on explaining the phenomena

rather than just making future predictions, Levi's was not only making predictions on how many baggy jeans it would be able to sell, they were understanding why this trend was occurring and making business decisions based on the insights. (*Canvas*, n.d.) Example of Levi's also highlights the data science pipeline, the steps in the data science pipeline are data ingestion, exploration, data wrangling, modeling, and deployment which Levi's has followed as it ingests large-scale data, analyses consumer behaviors, and makes informed business decisions. This aligns well with data science as a process rather than just a task. (*Applied Full Stack Data Science*, n.d.)

Additional data mining approaches that I might have taken are, I would do some sentimental analysis through social media platforms. As social media play a big role in today's styling and fashion trends, learn about comments and tags about baggy jeans, gather data about the preferences, positive as well as negative comments about baggy jeans which could provide more insights. Another one I could think of would be to look deep into regional data, I would look at where more baggy jeans are in demand and are in sales, for example, if baggy jeans are selling more in San Francisco in comparison to Boston, sending more stock to fulfill the demand in San Francisco would be ideal compared to equal stock supply in both cities.

The risks I can see what the analyst might have encountered or could encounter despite a successful data mining process are relying on past data too much which could keep focus on past successful models and lead to failure to adapt to a new market shift, let's say for example they could just focus on skinny jeans as they have sold much in the past despite baggy jeans making a comeback with a bang which leads to model overfitting. Another one is analysts focusing on their own beliefs which creates bias in data interpretation leading to misguided business decisions.

Article Link ⬈ **(http://www.wsj.com/articles/how-tech-helped-levis-ride-the-baggy-jeans-trend-f290721d)**

Thanks

References:

Applied Full stack Data science. (n.d.). Retrieved March 18, 2025, from https://endtoenddatascience.com/chapter2-defining-data-science

Bousquette, I. (2025, January 27). *How tech helped Levi's ride the "Baggy Jeans" trend*. mint. https://www.livemint.com/companies/news/how-tech-helped-levi-s-ride-the-baggy-jeans-trend-11737976806672.html

Redirecting. (n.d.). Retrieved March 18, 2025, from **https://canvas.park.edu/courses/85581/pages/unit-1-lecture?module_item_id=6163493 (https://canvas.park.edu/courses/85581/pages/unit-1-lecture?module_item_id=6163493)**

Reply | ✉

> **2 Replies (2)**