THE COMPUTERTECH
(An International Peer Review Journal)
VOLUME 11; ISSUE 1 (JAN-JUNE); (2025)
WEBSITE: THE COMPUTERTECH

Data Mining Curriculum – A Proposal for a Comprehensive and Industry-Oriented Academic Framework

Sai Dikshit Pasham[1], Y. P.
University of Illinois, Springfield, UNITED STATES

Abstract

The rapid growth of data across scientific, industrial, and social domains has created an increasing demand for professionals skilled in data mining and knowledge discovery. This proposal presents a structured and comprehensive data mining curriculum designed to equip students with both theoretical foundations and practical competencies required in modern data-driven environments. The proposed curriculum emphasizes core concepts such as data preprocessing, exploratory data analysis, classification, clustering, association rule mining, anomaly detection, and predictive modeling. In addition, it integrates essential topics including machine learning fundamentals, statistical analysis, big data technologies, data visualization, and ethical considerations in data usage. A strong focus is placed on hands-on learning through laboratory sessions, real-world datasets, case studies, and project-based assessments to bridge the gap between academic learning and industry requirements. The curriculum is designed to be modular and flexible, allowing adaptation across undergraduate and postgraduate programs while remaining aligned with evolving technological trends. Tools and platforms such as Python, R, SQL, and open-source data mining frameworks are incorporated to enhance practical exposure. Furthermore, the proposal highlights the importance of interdisciplinary learning by encouraging applications of data mining in healthcare, finance, business intelligence, cybersecurity, and scientific research. By fostering analytical thinking, problem-solving skills, and ethical awareness, this curriculum aims to prepare graduates for careers in data science, artificial intelligence, and advanced analytics, while also supporting research and innovation in the field of data mining.

Keywords: Data Mining; Curriculum; Industry-Oriented; Academic Framework

Introduction

Recent tremendous specialized advances in handling power, stockpiling limit, and between a network of computer innovation is making phenomenal amounts of computerized information. Information mining, the study of extricating valuable information from such immense information stores, has arisen as a youthful and interdisciplinary field in software engineering. Information mining strategies have been generally applied to industry, science, engineering, and government issues, and it is generally accepted that information mining will significantly affect our society. The developing agreement that data mining can bring genuine worth has prompted a blast of sought-after novel information-mining advances. For students prepared in information mining, understudies who comprehend information mining strategies can apply them to real issues and are ready to research, what's more, improvement of new information mining strategies. Courses in information mining have begun to spread everywhere in the world.

Because of this improvement of the field, the ACM SIGKDD Executive Committee has set up the ACM SIGKDD Curriculum Committee to plan an exemplary educational program for information mining that gives proposals for instructing the up and coming age of understudies in information mining. Given criticism from specialists, teachers, furthermore, understudies, we are persuaded that it is a significant undertaking to have a painstakingly planned, adroitly solid, in fact rich, and adjusted educational plan for this control. A complete and adjusted educational program will guarantee that the schooling in information mining sets a strong establishment for the solid development of the field, and it will advance precise preparation of understudies in software engineering, data sciences, and other related fields,

The Curriculum Committee comprises college educators and scientists who have con effectively-tributed to information mining examination and instruction, scientists and specialists from industry who have rich encounters in applying information mining innovation, and directors from government offices. This report is the principal draft from the Intensive Working Group of the Committee. We expect that this draft will be widely updated and checked. We anticipate ideas and proposals from the Council and the overall information mining examination, advancement, and application area.

In the first place, we diagram the rules that guided us in the choice of material in Section 2. We, at that point, give a short depiction of the requirements that we accept understudies of our proposed educational program to have in Section 3. Area 4 contains the center of this archive, our academic plan proposition.

CURRICULUM DESIGN PHILOSOPHY

Information mining is an interdisciplinary field at the convergence of artificial brainpower, AI, statistics, and information base frameworks. We accept that various teachers will stress multiple points in their courses. Consequently, we isolated this educational plan proposition into two sections. Foundations' initial segment contains essential material that we accept ought to be shrouded in any starting seminar on information mining. The second part, called Advanced Topics, is a far-reaching assortment of material that can be tested to finish an initial course or choices that can frame the reason for a high-level system in information mining.

We accept that the educating of data mining should focus on enduring logical standards and ideas of the field. Subsequently, instead of covering the last subtleties of the latest exploration, we planned the essential material to establish a robust framework that makes way for investigating further developed material.

The center undertaking in information mining is to extricate information from information; this information is caught in a human-reasonable design. The disclosure of construction in the statement is a diverse issue that incorporates the accompanying parts:

1. ***Database and Data Management Issues***: Where does the information dwell? How is it to be gotten? What types of inspecting are required? What information are conceivable? are proper? What are the ramifications of the information base or information distribution center construction and imperatives on information development and information readiness?
2. ***Data Preprocessing***: What are the necessary information changes before a picked calculation or class of measures can be applied to the information? What are the successful strategies for lessening the dimensionality of the data? How are missing

information things to be displayed? What changes appropriately encode deduced information on the issue?

3. ***The decision of Model and Statistical Inference Considerations***: What are the right choices to ensure legitimate factual surmising? What are fair approximations? What are the ramifications of the derivation strategies on the expected outcomes? How is the subsequent construction to be assessed? How is it approved?
4. ***Intriguing quality Metrics***: What makes the inferred structure fascinating or valuable? How do the specific information mining action objectives impact the selection of calculations or methods to be utilized?
5. ***Algorithmic Complexity Considerations***: What selection of calculations dependent on the size and dimensionality of information? What might be said about computational asset imperatives? Prerequisites on the exactness of coming about models? What are the versatility contemplations, and how could they be tended.
6. ***Post-processing of Discovered Structure***: How are the outcomes to be utilized? What are the prerequisites for use at the forecast time? What are the change necessities at model application time? How are differences in the information or hidden conveyances to be overseen?
7. ***Visualization and Understandability***: What imperatives on the found design from the point of view of understandability by people? What are effective perception procedures for the resulting structure? How could information be successfully pictured with regards to or with the guide of the found forms?
8. ***Maintenance, Updates, and Model Life Cycle Considerations***: When are models to be changed or refreshed? How should the models vary as the utility measurements in the application area change? How are the subsequent expectations or found construction incorporated with application area measurements and limitations?

The halfway rundown above exhibits that information mining includes numerous issues and numerous ideas that have genuinely been concentrated in confinement. This requires a solid inclusion of a broad scope of regions inside the proposed educational plan.

PREREQUISITES

Information mining is an expansive field that joins methods from various regions in software engineering and measurements.

Our model educational plan accepts that understudies have essential foundation information in the accompanying territories:

a. ***Database Systems***: Data models, question dialects, SQL, calculated data set plan, inquiry handling, furthermore, exchange handling.
b. ***Statistics***: Expectation, essential likelihood, dissemination, theory tests, ANOVA, and assessing an appropriation boundary.
c. ***Linear Algebra***: Vectors and grids, vector spaces, premise, lattice reversal, and addressing temperate conditions.
d. ***Algorithms and Data Structures***: We expect knowledge of fundamental information constructions and general maturity of understudies to comprehend calculations written in pseudo-code.

We accept that most software engineering seniors either have canvassed this material in past courses, can get missing material in self-study, or that the course teacher presented the disappeared material as essential.

COURSE TOPICS AND MODELS

Review that we apportioned our educational program into two sections: A seminar on Foundations and a seminar on Advanced Points. A standard 14-week one-semester initial conference on information mining (offered to one or the other senior undergrad or first-year graduate understudies) could cover every one of the units in Foundations and a chosen set of units from the Advanced Topics. A selected group of companies from the Advanced Topics can be canvassed in a second course.

1) FOUNDATIONS (COURSE I)

1. ***Presentation***. Fundamental data mining ideas, including inspiration, definition, the connections of data mining with data set frameworks, measurements, AI, various types of information stores on which information mining can be performed. Also, diverse sorts of examples and information to be mined, the idea of intriguing quality, and the latest things and improvements of information mining. The material can most likely be presented by showing a couple of contextual analyses.

    a. **Concepts of information mining**: inspiration, definition, the connections of information mining with data set frameworks, measurements, AI, and data recovery.
    b. **Knowledge revelation measure**: An outline of the Knowledge Discovery Process. Accentuation on the iterative and intelligent nature of the KDD Process.
    c. **Mining on various types of information**: social, value-based, object-social, heterogeneous, spatiotemporal, text, interactive media, Web, stream, versatile, etc.
    d. **Mining for various sorts of information**: characterization, relapse, bunching, regular patterns, discriminant, anomalies, etc.
    e. **Evaluation of information**: intriguing quality or nature of information, including exactness, utility (like help), and pertinence (like connection).
    f. **Applications of information mining**: market examination, logical and designing interaction investigation, bioinformatics, country security, etc.

2. ***Information Preprocessing***. This unit will cover the accompanying subjects: (1) why preprocess the information? (2) fundamental information cleaning strategies, (3) information coordination and change, and (4) information decrease techniques. In specific, the accompanying subjects will be covered.

    a) **Descriptive information rundown**: This unit covers essential procedures for summing up and depicting information. It will cover: (1) registering the proportions of focal inclination like mean; furthermore, mode, (2) registering

the proportions of information scattering, for example, quantiles, boxplots, differences, standard deviation, and anomalies, and (3) realistic showcase of essential measurable depictions, for example, histogram, disperse plot, boxplot, quantile-quantile plot, and nearby relapse bends.

b) **Data cleaning strategies**: Basic methods for dealing with missing qualities, boisterous information, and conflicting information, including normal binning, grouping, and relapse techniques for information cleaning.
c) **Data incorporation and change strategies**: This incorporates information smoothing, information aggregation, information speculation, standardization, trait (or highlight) development.
d) **Basic information decrease techniques**: It presents binning (histograms), inspecting, and information 3D square collection.
e) **Discretization and idea progression age**: It covers discretization and idea hierarchy age for numeric information (counting binning, bunching, histogram investigation) and straight-out information (programmed age of idea progressive systems).

3. ***Information Warehousing and OLAP for Data Mining.*** This unit presents the idea of an information distribution center and its related dimensional information model. It, at that point, presents a fundamental OLAP-style investigation on the solid information shape.

a) **Concept and design of information distribution center**
b) **The dimensional information model**: including measurements and measures; star blueprint, snowflake composition, and truth heavenly bodies; information block idea; idea chains of command in the 3D square.
c) **OLAP Operations**. OLAP activities in the multidimensional information model (drill-down, move up, cut up, rotate)

4. ***Affiliation, relationship, and regular pattern analysis***. This unit covers the ideas and procedures for affiliation, connection, and continued example investigation, including the accompanying subjects.

a) **Basic ideas**: continuous examples, affiliations, backing, and certainty of affiliation rules, relationship measure, other target capacities or measures, a typical application situation: market bushel investigation.
b) **Frequent example mining strategies**: (1) the Apriori calculation, (2) upgrades to Apriori, (3) digging for max-designs, shut examples, and top-k examples.
c) **Mining different sorts of common examples**: (1) staggered and multidimensional affiliation rules, (2) quantitative affiliation rules, and (3) connection examination.
d) **Applications of affiliation rules**: (1) Weblog investigation, (2) utilization of affiliation manages as classifiers

5. ***Grouping***. This unit covers the ideas and strategies for grouping examination, including the following points.

   i. **Basic ideas**: grouping
   ii. **Evaluation of grouping**: (1) assessment metric, (2) approval for model choice, (3) overfitting, (4) contrasting classifiers dependent on money-saving advantage and ROC bends
   iii. **Bayesian grouping**: (1) establishment: Bayes hypothesis, (2) Naive Bayesian arrangement methods
   iv. **Decision tree and choice principle acceptance**: (1) property choice and decrease, (2) essential hierarchical grouping tree acceptance diagram, (3) pre/post-pruning uninformative subtrees, (4) extraction of rules from characterization trees, (5) choice standard enlistment.
   v. **Linear models for characterization**: (1) straight discriminant investigation, (2) grouping by SVM (Backing Vector Machine) investigation
   vi. **Basic ideas of nonlinear characterization**: (1) neural organization, (2) SVM with nonlinear parts
   vii. **Classification by lethargic assessment**: (1) k-closest neighbor classifier: essential thought and mistake limits, (2) privately weighted learning
   viii. **Ensemble classifier**: Basic thoughts why troupe development helps, essentials of weighted democratic, packing, boosting.

6. ***Bunch and Outlier Analysis***. This unit covers the ideas and methods for bunch and anomaly examination, including the accompanying themes.

   i. Concept of group investigation
   ii. Types of information and for disparity calculation: Interval-scaled factors, paired factors, ostensible, ordinal, and proportion scaling factors, and factors of blended sorts.
   iii. An order of significant bunching techniques
   iv. Partition-based grouping: k-means and k-medoids calculations, and versatile apportioning strategies.
   v. Hierarchical grouping: agglomerative and troublesome progressive bunching strategies, small groups: coordinated and versatile various leveled bunching techniques.
   vi. Density-based grouping: idea of thickness-based bunching, versatile mining of bunching structures, collection dependent on thickness dispersion capacities.
   vii. Model-based grouping: (1) The EM Algorithm, (2) neural organization approach (SOM)
   viii. Outlier examination: Concepts and fundamental anomaly discovery techniques.

7. ***Mining Time-Series and Sequence Data***. This unit covers the strategies for mining time-arrangement, what's more, succession information, with the accompanying points.

   a) **Regression examination**: (1) primary and numerous direct relapse, (2) nonlinear relapse, (3) calculated relapse, (4) relapse trees, (5) relapse utilizing Support Vector Machine, (6) other relapse models.
   b) **Trend examination**: A factual methodology
   c) **Sequential example mining**: Mining various consecutive examples, successive example mining strategies, imperative based straight example mining, shut consecutive models, from consecutive models to incompletely requested samples.

8. **Text Mining and Web Mining**. This unit covers the strategies for mining text and Web information, counting the accompanying points.

   a) **Mining text data sets**: (1) Text information investigation and data recovery, (2) watchword based affiliation examination, (3) record order, (4) text bunching investigation
   b) **Mining the World-Wide Web**: (1) Mining the Web's connection constructions to recognize definitive Website page, (2) programmed characterization of Web records, (3) development of a diverse Web database, (4) mining informal communities, (5) Web asset disclosure, (6) Web use mining.

9. ***Visual Data Mining***. This unit covers the visual information mining procedures, including the accompanying themes.
   i. Data perception
   ii. Visualization of information mining results
   iii. Visual information mining: visual classifier, projection pursuits, class-protecting projections, visualizing class-construction of high-dimensional information, class visits
10. ***Data Mining***: Industry endeavors and social effects
    a) The social effect of information mining
    b) Data mining and protection
    c) Standardization endeavors
    d) Data mining framework items

2) ADVANCED TOPICS (COURSE II)

1. ***Progressed Data Preprocessing***. This unit will cover advanced information decrease techniques.

   a) **Advanced information decrease techniques**: (1) dimensionality decrease (highlight or characteristic sub-set decrease), (2) numerosity decrease (relapse,

histogram, grouping, inspecting, solitary esteem deterioration (SVD), and discretization), and (3) information pressure (lossless versus lossy pressure, Fourier and wavelet change, and head segment investigation).

2. ***Information Warehousing, OLAP, and Data Generalization***: This unit covers progressed material in information warehousing, OLAP, and information speculation

   a. **The multidimensional information model**
   b. **Implement information distribution centers**: information joining, ordering OLAP information (bitmap file), proficient preparing of OLAP questions, metadata storehouse, information stockroom back-end devices, and utilities.
   c. **Efficient calculation of information shapes**: arrangement of measures: distributive, logarithmic, and comprehensive measures, block calculation techniques, ice shelf 3D shapes, hierarchical and base up computation, figuring shut and surmised information blocks.
   d. **Other information speculation draws near**: Attribute-situated enlistment, mining class comparisons: segregating between various classes.
   e. **Exploration of information distribution center and information mining**: Discovery-driven investigation of information 3D squares, complex accumulation at various granularity, solid shape slope examination, from insightful online handling to online logical mining.

3. ***Progressed affiliation, relationship, and continuous example examination***.

   a. **Advanced regular example mining techniques**: (1) vertical arrangement mining, (2) design development calculation, (3) mining shut examples and max-designs
   b. **Constraint-based affiliation mining**: (1) rule-and-question guided affiliation mining, (2) hostile to monotonicity, monotonicity, conciseness in obliged mining, (3) convertible limitations.
   c. **Extensions and utilization of successive example mining**: (1) icy mass solid shape calculation, (2) installments and semantic information pressure, (3) successive example based characterization and bunch investigation

4. ***Progressed Classification***.

   a) **Bayesian conviction organizations**: strategies for (cutting edge) picking BBN construction and preparing Bayesian conviction organizations
   b) **Advanced choice tree development**: (1) improvements to fundamental order tree induction, (2) adaptable calculations for order tree enlistment, (3) incorporating information warehousing methods and grouping tree acceptance, (4) arrangement with part of the way named information

c) **Neural organization approach for grouping**: (1) a multi-facet feed-forward neural organization, (2) characterizing organization geography, (3) back-spread, (4) interpretability of grouping results.
d) **Kernel techniques**: (1) piece calculated relapse, (2) portion discriminant investigation, (3) progressed SVM bit strategies.
e) **Introduction to learning hypothesis**: PAC-learnability, exact, valid and underlying danger, VC- hypothesis.
f) **Ensemble development**: Weighted democratic, stowing, powerless student, boosting, AdaBoost
g) **Other arrangement strategies**: (1) case-based thinking, (2) hereditary calculations, (3) unpleasant set approach, (4) fluffy set methodology

5. ***Progressed group examination.***

a) **Grid-based grouping**: A measurable data lattice approach, bunching by wavelet examination, grouping high-dimensional space.
b) **Clustering high-dimensional information**: Subspace grouping, typical example-based bunching, grouping by wavelet examination.
c) **Advanced exception examination**: Statistical-based anomaly location, distance-based exception recognition, deviation-based exception identification, examination of nearby anomalies.
d) **Collaborative separating**

6. ***Progressed Time-Series and Sequential Data Mining***. This unit covers the high-level strategies for consecutive mining information, including the accompanying points.

a) **Similarity** search in a time-arrangement examination:
b) **Hidden Markov models**
c) **Periodicity examination**: Transformation-based methodology, mining fractional periodicity.
d) **Sequence division**: Hidden Markov model and Variable Markov model for succession segmentation.
e) **Sequence grouping and bunching**: (1) q-gram based techniques, watchword-based strategies; (2) (high request) Markov chain, covered up Markov model; (3) postfix tree, prospect addition tree, and probabilistic automata.

7. ***Mining Data Streams***: This unit covers the procedures for mining stream information, including the following themes.

a) What is stream information?
b) **Essential devices**: Chernoff limits, repository inspecting
c) Stream test tallying and continuous example examination
d) Classification of information streams

e) Clustering information streams
f) Online sensor information examination

8. ***Mining Spatial, Spatiotemporal, and Multimedia information***. This unit covers the strategies for mining spatial, spatiotemporal, and sight and sound knowledge, including the accompanying points.

   a. **Mining spatial and spatiotemporal data sets**: (1) Spatial information 3D equitable development and spatial OLAP, (2) spatial affiliation and co-area examination, (3) spatial bunching techniques, (4) spatial grouping and spatial pattern examination, (5) spatiotemporal information emulating, (6) mining moving articles and directions.
   b. **Mining media data sets**: (1) multidimensional investigation of sight and sound information, (2) comparability search insight and sound information, (3) arrangement and relapse examination of media information, (4) mining affiliation and relationship in media information, (5) grouping sight and sound information
   c. **Mining object information bases**: (1) multidimensional investigation of complex articles, (2) speculation on complex organized and semi-organized information, (3) strategy for mining complex object databases: conglomeration, estimate, and reformist refinement.

9. ***Mining Biological Data***: This unit covers the methods for organic mining information, including the following subjects.

   a. **Mining DNA, RNA, and proteins**: (1) Mining theme designs, (2) looking through homology in enormous information bases, (3) phylogenetic and utilitarian expectation.
   b. **Mining quality articulation information**: (1) grouping quality articulation, e.g., for quality administrative net-works, (2) ordering quality articulation, e.g., for sickness touchy quality revelation.
   c. Mining mass spectrometry information
   d. Mining and incorporating information from biomedical writing
   e. Mining between space affiliations

10. ***Text mining***. This module will cover work that applies known mining strategies to the content media, stressing the new issues.

   a) **Text portrayal**: Set-of-words, pack of-words, vector-space model; the issue of substantial crude dimensionality
   b) **Dimensionality decrease:** PCA, SVD, inert semantic ordering
   c) **Text bunching**: agglomerative, k-implies, EM; the impact of countless commotion measurements, incomplete oversight

d) **Feature determination** in high measurements
e) **Naive Bayes characterization**: Poor thickness gauges, minor degree Bayesian conviction network induction
f) **Discriminative learning**: most extreme entropy, strategic relapse, and backing vector learning
g) **Shallow phonetics**: Phrase location, grammatical feature labeling, named element extraction, word sense disambiguation

11. ***Hypertext and Web mining***. This module will cover work that is explicit to investigating hypermedia, i.e., including various leveled labeling dialects and hyperlinks related to message.

a) **Web demonstrating**: The Web as a developing, communitarian, libertarian interpersonal organization: total diagram construction of the Web, special connection connecting models and test approval
b) **Link mining and interpersonal organization investigation**: Links as support: PageRank and HITS calculations to distinguish legitimate Web pages; associations with bibliometry
c) **The PageRank calculation**: Integrating page substance and page format with connecting structure; touchy theme PageRanks; Google
d) **Mining by misusing text and connections**: Exploiting text and links for better grouping and characterization; bound together probabilistic models for text and connections
e) **Structured information extraction**: Information extraction, misusing markup construction to separate organized information from pages implied for human utilization
f) **Multidimensional Web data sets**: Automatic development of diverse Web database; finding elements and relations on the Web (WebKB)
g) **Exploration and asset revelation on the Web**: support learning, different methodologies
h) **Using the Web in mining and versatile Web locales**: Reorganizing Web destinations by mining log information

12. ***Information Mining Languages, Standards, and System Architectures***. This unit covers the issues identified with information mining dialects, guidelines, and framework structures, including the accompanying themes.

a) **Data mining natives**: what characterizes an information mining task? task-pertinent information, the sort of information to be mined, foundation information: idea chains of command, the client determined imperatives, intriguing quality measures, a show of found examples
b) **Data mining dialects**, UIs, and normalization endeavors
c) **Architectures of information mining frameworks**

13. ***Information Mining Applications***. This unit covers the issues identified with space explicit information mining applications, including the accompanying subjects. (Note: Some of these subjects, if concrete and excellent materials are accessible, ought to go into the Foundations part as contextual investigations.)

    a. Data mining for monetary information examination
    b. Data mining for the retail business
    c. Data mining for the telecom business
    d. Data mining for interruption recognition
    e. Data mining is logical and measurable applications
    f. Data mining in programming and PC framework investigation

14. ***Information Mining and Society***. This unit covers the issues identified with the social effects of information mining, counting the accompanying subjects.

    a. Social effects of information mining
    b. Data mining versus information security and protection
    c. Privacy-saving information mining

15. ***Patterns in Data Mining***. This unit covers the significant ways in information mining, including the accompanying Points.

    (a) Setting vital hypothetical establishments for information mining
    (b) Mining somewhere down in explicit applications
    (c) Ubiquitous and invisible information mining
    (d) Integrated information and data frameworks

DIFFERENT COURSE MODULES & EDUCATIONAL GOALS

The course can be educated in various fields, like software engineering, business, and insights. The system can also be familiarized with multiple accentuations, data sets, data frameworks, and AI. We ought not to expect the material to be canvassed full range with comparative accentuation. We intend to embed a few modules dependent on the inputs of educators who have shown materials in explicit fields.

LABORATORIES AND EXERCISES

Research facilities and activities offer understudies a chance to complete investigations that represent subjects in a practical setting and simultaneously gain proficiency with the points of interest of the product utilized. Understudies may likewise be relegated to chip away at projects too huge ever to be finished during a solitary class period. Research facilities can give time to autonomous task work and programming tasks, revealing like that done in different points in software engineering.

The lab tasks can be ordered into a few classifications, and more inventive thoughts and ideas are energized.

1. Figure out how to utilize information mining frameworks by using some information mining and information warehousing virtual products. Run of the mill such programming projects may incorporate Microsoft SQLServer 2005 (Analysis supervisor), Oracle 10g (information mining part), IBM Intelligent-Miner, and measurements examination programming devices.

2. Carry out some information mining capacities, including affiliation mining, order, bunching, sequential design mining, text-mining, Web mining, bio-mining, spatial information mining bundles. On the other hand, some open or incompletely open-source information mining frameworks, like Weka, IlliMine, etc., can be utilized for information mining calculation augmentation and information mining application investigation.

3. Execution, refinement, and execution examination of a few distinctive information mining strategies.

4. Proposition, execution, and testing of new information mining calculations and capacities.

Utilizing some example informational indexes to execute and test information mining capacities, like KDD CUP information sets, UC-Irvine Machine Learning/KDD Repository, DBLP information base, and other chose Web information sets.

Conclusion

The proposed Data Mining Curriculum offers a forward-looking academic framework that aligns educational objectives with industry demands and technological advancements. By combining strong theoretical foundations with extensive practical training, the curriculum ensures that learners acquire both conceptual understanding and applied skills. Its modular design, interdisciplinary focus, and emphasis on ethical data practices make it adaptable and sustainable for long-term implementation. Ultimately, this curriculum proposal aims to produce competent, responsible, and industry-ready data mining professionals capable of addressing complex data challenges in a rapidly evolving digital world

References

[1] Feng, P., Bi, Z., Wen, Y., Pan, X., Peng, B., Liu, M., ... & Jiang, Z. (2024). Deep Learning and Machine Learning, Advancing Big Data Analytics and Management: Unveiling AI's Potential Through Tools, Techniques, and Applications. arXiv preprint arXiv:2410.01268.

[2] Cherukuri, R., & Parimi, S. K. (2025). AI-Enabled Architecture for Large-Scale Multi-Tenant Systems. International Journal of Modern Computing, 8(1), 1-10.

[3] Jaladi, D. S., & Vutla, S. (2017). Harnessing the Potential of Artificial Intelligence and Big Data in Healthcare. The Computertech, 31-39.

[4] Cherukuri, R., & Putchakayala, R. (2021). Frontend-Driven Metadata Governance: A Full-Stack Architecture for High-Quality Analytics and Privacy Assurance. International Journal of Emerging Research in Engineering and Technology, 2(3), 95-108.

[5] Gudepu, B. K., & Jaladi, D. S. (2018a). The Role of Data Profiling in Improving Data Quality. The Computertech, 21-26.

[6] Putchakayala, R., & Cherukuri, R. (2022). AI-Enabled Policy-Driven Web Governance: A Full-Stack Java Framework for Privacy-Preserving Digital Ecosystems. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 3(1), 114-123.

[7] Cohen, M. (2022). Rethinking Thinking: Problem Solving from Sun Tzu to Google. Andrews UK Limited.

[8] Putchakayala, R., & Cherukuri, R. (2024). AI-Enhanced Event Tracking: A Collaborative Full-Stack Model for Tag Intelligence and Real-Time Data Validation. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(2), 130-143.

[9] Cherukuri, R., & Putchakayala, R. (2022). Cognitive Governance for Web-Scale Systems: Hybrid AI Models for Privacy, Integrity, and Transparency in Full-Stack Applications. International Journal of AI, BigData, Computational and Management Studies, 3(4), 93-105.

[10] Jaladi, D. S., & Vutla, S. (2018a). An Analysis of Big Data Analytics in Relation to Artificial Intelligence and Business Intelligence. The Computertech, 37-46.

[11] Cherukuri, R., & Yarram, V. K. (2023). AI-Orchestrated Frontend Systems: Neural Rendering and LLM-Augmented Engineering for Adaptive, High-Performance Web Applications. International Journal of Emerging Research in Engineering and Technology, 4(3), 107-114.

[12] Jaladi, D. S., & Vutla, S. (2018b). The Use of AI and Big Data in Health Care. The Computertech, 45-53.

[13] Cherukuri, R., & Yarram, V. K. (2024). From Intelligent Automation to Agentic AI: Engineering the Next Generation of Enterprise Systems. International Journal of Emerging Research in Engineering and Technology, 5(4), 142-152.

[14] Gudepu, B. K., & Jaladi, D. S. (2018b). The Role of Data Quality Scorecards in Measuring Business Success. The Computertech, 29-36.

[15] Parimi, S. K., & Cherukuri, R. (2024). Proactive AI Systems: Engineering Intelligent Platforms that Sense, Predict, and Act. International Journal of Emerging Trends in Computer Science and Information Technology, 5(3), 122-130.

[16] Jaladi, D. S., & Vutla, S. (2024b). The Role of Artificial Intelligence in Modern Medicine. The Metascience, 2(4), 96-106.

[17] Parimi, S. K., & Yallavula, R. (2023). Enterprise Risk Intelligence: Machine Learning Models for Predicting Compliance, Fraud, and Operational Failures. International Journal of Emerging Trends in Computer Science and Information Technology, 4(2), 173-181.

[18] Xiong, Z., Liang, Z., Liu, X., Feldmann, M., & Li, J. (2023). Steel-UHPC composite dowels' pull-out performance studies using machine learning algorithms. Steel and Composite Structures, 48(5), 531-545.

[19] Parimi, S. K., & Yallavula, R. (2025). Generative AI for Enterprise Trust: A Governance-Aligned Framework for Safe and Transparent Automation at Global Scale. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 6(1), 218-225.

[20] Himeur, Y., Varlamis, I., Kheddar, H., Amira, A., Atalla, S., Singh, Y., ... & Mansoor, W. (2023). Federated learning for computer vision. arXiv preprint arXiv:2308.13558.

[21] Yarram, V. K., & Parimi, S. K. (2024). The Next Frontier of Enterprise Transformation: A Comprehensive Analysis of Generative AI as a Catalyst for Organizational Modernization, Intelligent Automation, and Large-Scale Knowledge Acceleration Across Global Digital Ecosystems. The Metascience, 2(2), 97-106.

[22] Jaladi, D. S., & Vutla, S. (2024a). Machine Learning Techniques for Analyzing Large-Scale Patient Databases. International Journal of Modern Computing, 7(1), 181-198.

[23] Yallavula, R., & Putchakayala, R. (2022). A Data Governance and Analytics-Enhanced Approach to Mitigating Cyber Threats in NoSQL Database Systems. International Journal of Emerging Trends in Computer Science and Information Technology, 3(3), 90-100.

[24] Jaladi, D. S., & Vutla, S. (2020b). Machine Learning Demystified: Concepts, Algorithms, and Use Cases. The Computertech, 1-12.

[25] Yallavula, R., & Putchakayala, R. (2024). AI for Data Governance Analysts: A Practical Framework for Transforming Manual Controls into Automated Governance Pipelines. International Journal of AI, BigData, Computational and Management Studies, 5(1), 167-177.

[26] Jaladi, D. S., & Vutla, S. (2021b). Quantum AI: Accomplishments and Obstacles in the Convergence of Quantum Computing and Artificial Intelligence. International Journal of Modern Computing, 4(1), 86-95.

[27] Yallavula, R., & Yarram, V. K. (2021). An AI Framework for Monitoring Rule Changes in Highly Volatile Compliance Environments. The Computertech, 39-53.

[28] Gudepu, B. K., & Jaladi, D. S. (2022b). Why Real-Time Data Discovery is a Game Changer for Enterprises. International Journal of Acta Informatica, 1(1), 164-175.

[29] Jaladi, D. S., & Vutla, S. (2022a). Artificial Intelligence's Influence on Design: A New Era of Creative Collaboration. International Journal of Acta Informatica, 1(1), 188-198.

[30] Yarram, V. K., & Cherukuri, R. (2023). From Data to Decisions: Architecting High-Performance AI Platforms for Fortune 500 Ecosystems. The Metascience, 1(1), 306-324.

[31] Wu, X., Jiang, L., Wang, P. S., Liu, Z., Liu, X., Qiao, Y., ... & Zhao, H. (2024). Point transformer v3: Simpler faster stronger. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4840-4851).

[32] Yallavula, R., & Putchakayala, R. (2023). Governance-of-Things (GoT): A Next-Generation Framework for Ethical, Intelligent, and Autonomous Web Data Acquisition. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 4(4), 111-120.

[33] Gudepu, B. K., Jaladi, D. S., & Gellago, O. (2023). How Data Catalogs are Transforming Enterprise Data Governance: A Systematic Literature Review. The Metascience, 1(1), 249-264.

[34] Yarram, V. K., & Yallavula, R. (2022). Adaptive Machine Learning Driven Compliance Scoring Models for Automated Risk Detection, Quality Validation of AI-Generated Content in Regulated Industries. International Journal of Emerging Research in Engineering and Technology, 3(1), 116-126.

[35] Parimi, S. K., & Yarram, V. K. (2022). AI-First Enterprise Architecture: Designing Intelligent Systems for a Global Scale. The Computertech, 1-18.

[36] Koç, B., & Çetintaş, D. (2024). Diagnosis Enhancement Using Deep Learning: YOLOv11-YOLOv8 for Alzheimer's and Parkinson's Disease Detection. UAKK 2024, 13.

[37] Kshetri, N., Hughes, L., louise Slade, E., Jeyaraj, A., kumar Kar, A., Koohang, A., ... & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642.

[38] Putchakayala, R., & Parimi, S. K. (2023). AI-Optimized Full-Stack Governance A Unified Model for Secure Data Flows and Real-Time Intelligence. International Journal of Modern Computing, 6(1), 104-112.

[39] Jaladi, D. S., & Vutla, S. (2019a). Deploying Breiman's Random Forest Algorithm in Machine Learning. The Computertech, 45-57

[40] Putchakayala, R., & Yallavula, R. (2024). AI-Driven Federated Data Governance: Building Trustworthy and Sustainable Digital Ecosystems. International Journal of Modern Computing, 7(1), 219-227.

[41] Yarram, V. K., & Cherukuri, R. (2025). User Experience Patterns for Front-End Integration of Retrieval-Augmented Generation in Enterprise Platforms. International Journal of Emerging Trends in Computer Science and Information Technology, 6(2), 87-94.

[42] Williams, R., Ali, S., Devasia, N., DiPaola, D., Hong, J., Kaputsos, S. P., ... & Breazeal, C. (2023). AI+ ethics curricula for middle school youth: Lessons learned from three project-based curricula. International Journal of Artificial Intelligence in Education, 33(2), 325-383.

[43] Yarram, V. K., & Parimi, S. K. (2021). Design and Implementation of a Responsible, Explainable, and Compliance-Driven AI Architecture for Enterprise-Scale Content Management Systems Integrating Generative Models, Retrieval Pipelines, and Real-Time Governance Controls. International Journal of Modern Computing, 4(1), 96-110.

[44] Jaladi, D. S., & Vutla, S. (2019b). Revolutionizing Healthcare Through Quantum Computing: Insights and Future Directions. International Journal of Modern Computing, 2(1), 60-83.

[45] Putchakayala, R., & Yallavula, R. (2025). The Intelligent Governance Core: A Multi-Layer AI Framework for Predictive Compliance and Autonomous Digital Analytics. International Journal of AI, BigData, Computational and Management Studies, 6(1), 171-179.

[46] Lurie-Luke, E. (2024). Alternative protein sources: science powered startups to fuel food innovation. Nature Communications, 15(1), 4425.

[47] Putchakayala, R., & Yarram, V. K. (2025). Adaptive Trust Engineering: AI-Driven Full-Stack Mechanisms for Privacy, Compliance, and Data Quality. The Metascience, 3(2), 1-21.

[48] Yallavula, R., & Parimi, S. K. (2022). Bridging Data, Intelligence, and Trust the Future of Computational Systems and Ethical AI. International Journal of Modern Computing, 5(1), 119-129.

[49] Jaladi, D. S., & Vutla, S. (2020a). Leveraging Data Mining to Innovate Agricultural Applications. International Journal of Modern Computing, 3(1), 34-46.

[50] Parimi, S. K., & Yallavula, R. (2021). Data-Governed Autonomous Decisioning: AI Models for Real-Time Optimization of Enterprise Financial Journeys. International Journal of Emerging Trends in Computer Science and Information Technology, 2(1), 89-102.