

Extracted Slides Data

Slide 1

Introducing , Prezi AI. Your new presentation assistant. Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker, than ever before. Search presentations... tii (cametn emaanty What is data mining? Humans Machines rotstesl aed eomputinl Wehner dean br arch amido annua, secon boned sare rms ch esata. dt chores data ae Human-Machine combines, What telnformation? norman ta at een pected ann oramsaue away nt maar RessLunderaanding ane nine a eon a sod daciler-eing prea sivng, tammuniation a sed maie tamed decionees pseiens Toasting he a wi

The screenshot shows a Prezi presentation slide. The slide features a central image of a laptop on a wooden desk. On the laptop screen, there is a diagram with a central black box containing the text 'ENABLING TECHNOLOGIES FOR DATA SCIENCE-I DATA MINING'. Surrounding this central box are several colorful sticky notes with the following text: 'DATA PREPROCESSING' (light blue), 'ASSOCIATIVE CLASSIFICATION' (pink), 'PREDICTION' (light blue), 'WEB SCRAPING' (green), 'TASK 7' (yellow), 'KDD PROCESS IN DATA MINING' (orange), 'DATA MINING' (yellow), 'TASK 8' (orange), and 'TASK 9' (light blue). The Prezi logo is visible in the bottom left corner of the slide image. Below the slide image, there is a white box containing the title 'Enabling Technologies for Data Science-I DATA MINING', a view count of 895, and three buttons: 'Make a copy', 'Create your own with AI', and 'Download PDF'.

Slide 2

Introducing , Prezi AI. Your new presentation assistant. Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker, than ever before. Search presentations... tii (cametn emaanty What is data mining? Humans Machines rotstesl aed eomputinl Wehner dean br arch amido annua, secon boned sare rms ch esata. dt chores data ae Human-Machine combines, What telnformation? norman ta at een pected ann oramsaue away nt maar RessLunderaanding ane nine a eon a sod daciler-eing prea sivng, tammuniation a sed maie tamed decionees pseiens Toasting he a wi

asectinsschtcutig, esietn reson as serene ining Sdnestan tnt canna Fu cane paraig
 ataingeanterandiecentty Why datais important?mereagnaete ard urge matategcargaln while
 nasttcaatear beuatatnatynac ff Dua hzaineabe eae eens,rao tr asasean cues persnatnes
 tment pms. seer insl cans by igh azn for anders,Howermidets tig dois citer aman parity
 whenitinaie pemarator ff D8 dpe neo stutter,sess dela sipartat i nr ates ing ended cic endwih
 spear OH Brechna Ino raeFeteuans nen protect the cy frien pve use of den Dts tos on wretard
 wns ad hemeWhat is Data?(Aeartg te Ove Ors ait plans of oreo scaly farm om spel wa
 sensoDeteanie mesa ela cpt andarle,wttespant cen dale rg oP,logan orth anata ne ta unpre
 may 3 alecn fumes orarash tare and ona eases fmt bored ra canren ntl stuart dt ety ens
 Dopey cermonl crn soge ardpares the roost dt mona age eld ale ba conrad he ow ot
 esabseaeFees aeneenencie erovelpioecnnen iokatemesiearoreanerety he abueaonef cute
 nigra Datta bw ameEnabling Technologies for Data Science-I dATA mININGos MOOE)Make
 a copy, Create your own with AI Download PDF)

Introducing :+ Prezi AI. Your new presentation assistant.
 Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker than ever before. [See how →](#)

Prezi Search presentations... [Log in](#) [Get started](#)

TRENDING SEARCHES [Timeline](#) [Lesson Plan](#) [Great Gatsby](#) [Template](#) [Brainstorm](#)

What is data mining?
 Data mining is the process of extracting knowledge or insights from large amounts of data using various statistical and computational techniques. The data can be structured, semi-structured or unstructured, and can be stored in various forms such as databases, data warehouses, and data lakes.
 The primary goal of data mining is to discover hidden patterns and relationships in the data that can be used to make informed decisions or predictions. This involves exploring the data using various techniques such as clustering, classification, regression analysis, association rule mining, and anomaly detection.
 Data mining has a wide range of applications across various industries, including marketing, finance, healthcare, and telecommunications. For example, in marketing, data mining can be used to identify customer segments and target marketing campaigns, while in healthcare, it can be used to identify risk factors for diseases and develop personalized treatment plans.
 However, data mining also raises ethical and privacy concerns, particularly when it involves personal or sensitive data. It's important to ensure that data mining is conducted ethically and with appropriate safeguards in place to protect the privacy of individuals and prevent misuse of their data.

What is Information?
 Information is data that has been processed, organized, or structured in a way that makes it meaningful, valuable and useful. It is data that has been given context, relevance and purpose. It gives knowledge, understanding and insights that can be used for decision-making, problem-solving, communication and various other purposes.

Why data is important ?
 • Data helps in make better decisions.
 • Data helps in solve problems by finding the reason for underperformance.
 • Data helps one to evaluate the performance.
 • Data helps one improve processes.
 • Data helps one understand consumers and the market.

What is Data?
 According to the Oxford "Data is distinct pieces of information, usually formatted in a special way". Data can be measured, collected, reported, and analyzed, whereupon it is often visualized using graphs, images, or other analysis tools. Raw data ("unprocessed data") may be a collection of numbers or characters before it's been "cleaned" and corrected by researchers. It must be corrected so that we can remove outliers, instruments, or data entry errors. Data processing commonly occurs in stages, and therefore the "processed data" from one stage could also be considered the "raw data" of subsequent stages. Field data is data that's collected in an uncontrolled "in situ" environment. Experimental data is the data generated from the observation of scientific investigations. Data can be generated by:

- Humans
- Machines
- Human-Machine combines.

DATA MINING

Enabling Technologies for Data Science-I dATA mINING

895 [❤](#) [🔗](#) [📌](#) [🔧](#)

[Make a copy](#) [Create your own with AI](#) [Download PDF](#)

Slide 3

Introducing , Prezi AI. Your new presentation assistantRefineevenhance,and tailor your
 content source relevant imagesyand edit visuals quicker,than ever before.)iT RENDING
 SEARCHESCateceries of Dota ant wel ranmeGoneratsiae neomnes cage cng wing Se lgmen

hued Elaaresaacrba atirroamatnthe ff 1 fsbo eat sete Pek Atom tetaroeandmas Ste 1 Sauna Beercds no rer snes enna ber om Bote aes opnaer archlee atone peeerg omeTypes of Data S.reionctec move ma ben acer ve uig wthi diferente teGenerally data can be classified into two parts: Femnnncinnvtmoaberreprton cveest ara fete en hepenny Sweet Scan era ceninons mriEnabling Technologies for Data Science-I dATA mININGos MOOE)Make a copy, Create your own with AI Download PDF)

Introducing **Prezi AI**. Your new presentation assistant. Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker than ever before. [See how →](#)

Prezi Search presentations... [Log in](#) [Get started](#)

TRENDING SEARCHES [Timeline](#) [Lesson Plan](#) [Great Gatsby](#) [Template](#) [Brainstorm](#)

Categories of Data

Data can be categorized into two main parts -

- Structured Data:** This type of data is organized data into specific format, making it easy to search, analyze and process. Structured data is found in a relational databases that includes information like numbers, data and categories.
- Unstructured Data:** Unstructured data does not conform to a specific structure or format. It may include some text documents, images, videos, and other data that is not easily organized or analyzed without additional processing.

Types of Data

Generally data can be classified into two parts:

- Categorical Data:** In categorical data we see the data which have a defined category, for example:
 - Marital Status
 - Political Party
 - Eye colour
- Numerical Data:** Numerical data can further be classified into two categories:
 - Discrete Data:** Discrete data contains the data which have discrete numerical values for example Number of Children, Defects per Hour etc.
 - Continuous Data:** Continuous data contains the data which have continuous numerical values for example Weight, Voltage etc.
- Nominal Scale:** A nominal scale classifies data into several distinct categories in which no ranking criteria is implied. For example Gender, Marital Status.
- Ordinal Scale:** An ordinal scale classifies data into distinct categories during which ranking is implied. For example:
 - Faculty rank - Professor, Associate Professor, Assistant Professor
 - Students grade - A, B, C, D, E, F
- Interval scale:** An interval scale may be an ordered scale during which the difference between measurements is a meaningful quantity but the measurements don't have a true zero point. For example:
 - Temperature in Fahrenheit and Celsius.
 - Years
- Ratio scale:** A ratio scale may be an ordered scale during which the difference between the measurements is a meaningful quantity and therefore the measurements have a true zero point. Hence, we can perform arithmetic operations on ratio scale data. For example : Weight, Age, Salary etc.

101 2018 Chend T. Jom

Enabling Technologies for Data Science-I dATA mINING

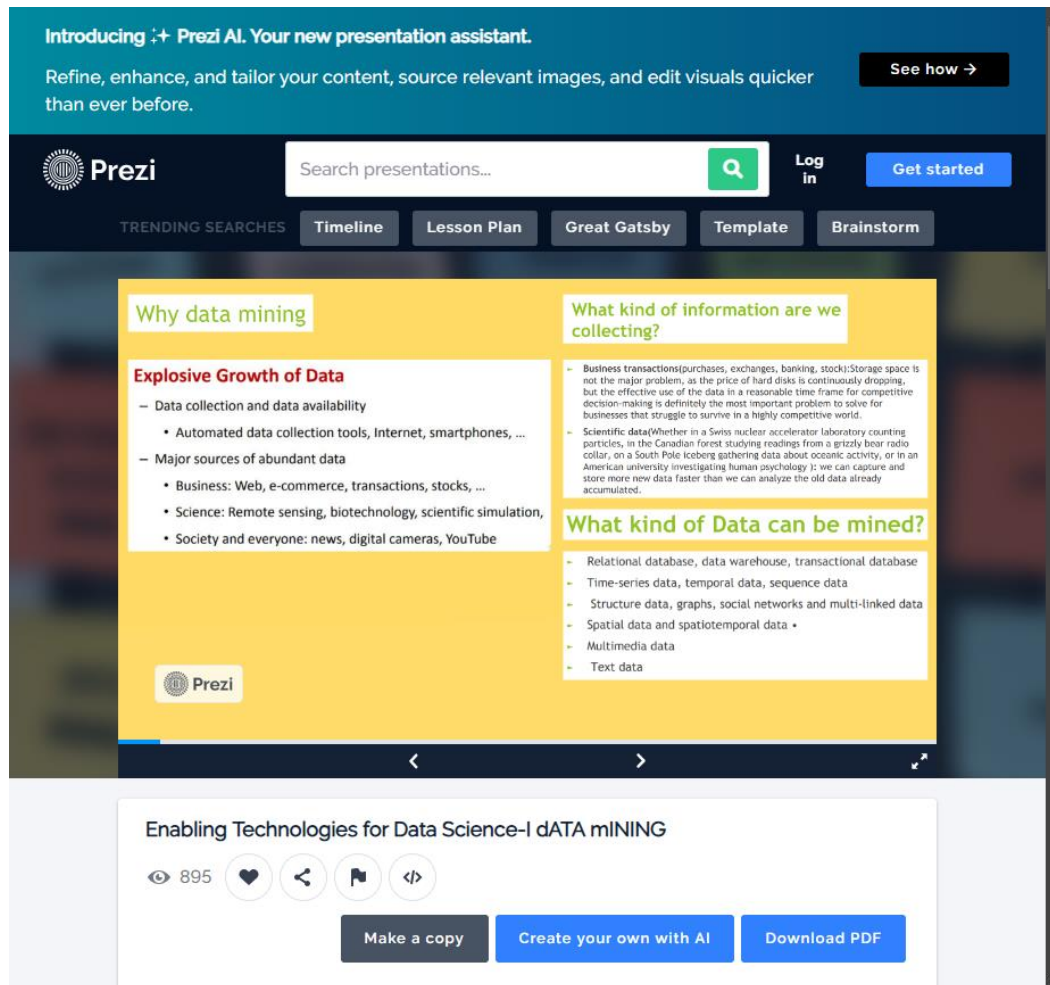
895 [❤](#) [🔗](#) [📌](#) [🔧](#)

[Make a copy](#) [Create your own with AI](#) [Download PDF](#)

Slide 4

Introducing , Prezi AI. Your new presentation assistantRefine, enhance, and tailor your content,source relevant images, and edit visuals quicker,than ever before.)Search presentations...Explosive Growth of Data Meite muppet ube ceattad ance Satay aepsae,iit he ete wea the data ta reonabe tine rome ar eomgerive- Scceonemaking ditch mom inprangrtom 4 ekData collection and data avallaity isomer thar svugletosunbe ina gy compte voi Automated data collection tools, internet, smartphones, Sede daira Sei rule cele abort cingDareles the Caran fest sca easng rom a ray bara- r Cela, South Fle cer sae data abot ocean Sty, or aMajor sources of abundant data Aanevcanuniversiyvestigating funan psychology Ye we ean eager andare more none font Bone oneal Be dete cayscore Business: Web, e-commerce, transactions, stocks, Science: Remote sensing, biotechnology, scientific

simulation, Society and everyone: news, digital cameras, YouTube
 Relational database, data warehouse, transactional database
 Time-series data, temporal data, sequence data
 Structure data, graphs, social networks and multi-linked data
 Spatial data and spatiotemporal data
 Multimedia data
 Text data
 Enabling Technologies for Data Science-I
 DATA MINING
 MOOE)
 Make a copy, Create your own with AI
 Download PDF)



Slide 5

Introducing , Prezi AI. Your new presentation assistant
 Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker, than ever before.)
 TRENDING SEARCHES
 Whats the Data Processing Pipeline?
 The data processing pipeline refers to the iterative sequence of transformations applied to raw data to generate meaningful insights. It can be viewed as a pipeline with data stages:
 1. Data Acquisition: This stage encompasses the methods used to collect raw data from various sources. This could involve sensor readings, scraping web data, or gathering information through surveys and application logs.
 2. Data Preparation: Raw data is often messy and requires cleaning and pre-processing before analysis. This stage involves tasks like identifying and handling missing values, correcting inconsistencies, formatting data into a consistent structure, and potentially removing outliers.
 3. Data Inputs: The pre-processed data is loaded into a system suitable for

further processing and analysis, This often involves converting the data into a machine-readable format and storing it in a database or data warehouse,4. Data Processing Here, the data undergoes various manipulations and transformations to extract valuable information. This may include aggregation, filtering, sorting, feature engineering (creating new features from existing ones) and applying machine learning algorithms to uncover patterns and relationships,5, Data Output: The transformed data is then analyzed using various techniques to generate insights and knowledge. This could involve statistical analysis, visualization techniques, or building predictive models,6. Data Storage: The processed data and the generated outputs are stored in a secure and accessible reference, or feeding into further analysis cycles,Enabling Technologies for Data Science-I dATA MININGos MOOE)Make a copy, Create your own with AI Download PDF)

Introducing :+ Prezi AI. Your new presentation assistant.

Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker than ever before.

See how →

Search presentations...

Log in

Get started

TRENDING SEARCHES

Timeline Lesson Plan Great Gatsby Template Brainstorm

What is the Data Processing Cycle?

The data processing cycle refers to the iterative sequence of transformations applied to raw data to generate meaningful insights. It can be viewed as a pipeline with distinct stages:

- 1. Data Acquisition:** This stage encompasses the methods used to collect raw data from various sources. This could involve sensor readings, scraping web data, or gathering information through surveys and application logs.
- 2. Data Preparation:** Raw data is inherently messy and requires cleaning and pre-processing before analysis. This stage involves tasks like identifying and handling missing values, correcting inconsistencies, formatting data into a consistent structure, and potentially removing outliers.
- 3. Data Input:** The pre-processed data is loaded into a system suitable for further processing and analysis. This often involves converting the data into a machine-readable format and storing it in a database or data warehouse.
- 4. Data Processing:** Here, the data undergoes various manipulations and transformations to extract valuable information. This may include aggregation, filtering, sorting, feature engineering (creating new features from existing ones), and applying machine learning algorithms to uncover patterns and relationships.
- 5. Data Output:** The transformed data is then analyzed using various techniques to generate insights and knowledge. This could involve statistical analysis, visualization techniques, or building predictive models.
- 6. Data Storage:** The processed data and the generated outputs are stored in a secure and accessible format for reference, or feeding into further analysis cycles.

STAGES OF DATA PROCESSING CYCLE

Input Stage	Processing Stage	Output Stage	Storage Stage
Social media	Machine learning	Reporting	Storing data
Surveys	Feature engineering	Visualization	Archiving data
Log files	Transformation	Exporting data	Retrieving data
Data acquisition	Preprocessing	Analysis	Storage

Enabling Technologies for Data Science-I dATA MINING

895

Make a copy Create your own with AI Download PDF

Slide 6

Introducing , Prezi AI. Your new presentation assistant(Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker,than ever before.)Search presentations...Data Mining ProcedureDineDicnarpsons tank wigs Som mara enrebiow pckredetaometiontingesi pedcon and deci CID fen eed Up Da eg ttn Sateen Toblrtsitacgmeatigt!pateneand tering dca opmionion wars, Dats Pinaaoe edeay

edNedinatanig aitafece Wing emo nyepnEan snes ome ON,Aer pen oss eter eenTen atien
 netanan na hpea Dama Deamepecoinranee ONAN,seth Moe nes a mca aFavor mt Foran Sed
 a Orem aPetar pci ai ine etal datondononeRasodiasactndaiteer Cpratshapeemaritaepers
 SercoTakingetiordeckenton Saget mt apecotterLeyesinetgreemncaein Pr oneme nen
 ronaniG Prez onEnabling Technologies for Data Science-I dATA mININGos MOOE)Make a
 copy, Create your own with AI Download PDF)

Introducing **Prezi AI**. Your new presentation assistant.
 Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker than ever before. [See how →](#)

Prezi Search presentations... [Log in](#) [Get started](#)

TRENDING SEARCHES [Timeline](#) [Lesson Plan](#) [Great Gatsby](#) [Template](#) [Brainstorm](#)

Data Mining vs. DBMS

Aspect	Data Mining	DBMS
Purpose	Discover patterns, trends, and insights from data.	Store, manage, and retrieve structured data.
Functionality	Analysis, prediction, and decision-making.	CRUD (Create, Read, Update, Delete) operations.
Focus	Extracting meaningful patterns and insights.	Ensuring data organization, accessibility, and security.
Techniques	Machine learning, statistical analysis, clustering, regression.	Indexing, normalization, query optimization.
Tools/Platforms	WEKA, RapidMiner, Python (scikit-learn, pandas).	MySQL, PostgreSQL, Oracle, SQL Server, MongoDB.
Input	Raw or preprocessed data from various sources.	Structured data (tables, rows, columns).
Output	Patterns, rules, predictions, visualizations.	Retrieved or manipulated data based on queries.
Dependency	Relies on data stored in DBMS for analysis.	Operates independently but serves as a data source for mining.
Example Scenario	Predicting customer churn based on usage patterns.	Storing customer records and usage data.
Use Cases	Business intelligence, fraud detection, market analysis.	Payroll systems, inventory management, CRM.

Data Mining Procedure

Enabling Technologies for Data Science-I dATA mINING


895 [❤️](#) [🔗](#) [📌](#) [🔧](#)

[Make a copy](#) [Create your own with AI](#) [Download PDF](#)

Slide 7

Introducing , Prezi AI. Your new presentation assistantRefine, enhance, and tailor your
 content, source relevant images, and edit visuals quicker,than ever before.)iT RENDING
 SEARCHESHow Do We Analyza Data? eerandnm taData anal constvtes the man step of date
 cea Inwhich we discover nowadge ond meaningful sys qe enscsmytane secon wneste ice os
 seoiem ames healrate Rs he eacueg ees wea teh ota sand eon tr ase gets. neers emi dnconson
 te nana graurcnstseeeHess a reais a tke specs voles Hens nbc othe key aspeetsimotrede ff
 wonenences eoversnog aera ore meen gear oa1. Defice Goals and Qeuections2 Choose the
 Right Techniquesual ears mort ct ta ing arnt choreSree ens He ae socom melon ne
 sonecrsnon tes Sa Anstey ln xi re tn an sr iontrear rego rin nd pop ta Arango anata caeNaan
 nn agatha en rl it nt pettyteen coup tee sagas pn on eneprint acon tacoBa hagas nt cs of gone

bt and oc alrina trot ta Ties atten as leargen custobas ventana gat an ssh eigen bit tf,i
pgsespeere ota denier enw ene tobe cen rsrrewrEnabling Technologies for Data Science-I
dATA mININGos MOOE)Make a copy, Create your own with AI Download PDF)




Introducing

Prezi AI. Your new presentation assistant.

Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker than ever before.

See how →



Search presentations...

Q

Log in

Get started

TRENDING SEARCHES

Timeline

Lesson Plan

Great Gatsby

Template

Brainstorm

How Do We Analyze Data?

Data analysis constitutes the main step of data cycle in which we discover knowledge and meaningful information from raw data. It's like reaching deep into the hands of a sand pile, looking for those gems. Here's a breakdown of the key aspects involved. Here's a breakdown of the key aspects involved:

1. Define Goals and Questions

To begin with, analyze what you need the data for, or in other words, determine your goals. Are you trying to do seasonal line ups, determine customer behavior or make forecasting? Clearly defined goals, indeed practical analysis techniques will be the key factor to ensure alignment to them.

2. Choose the Right Techniques

Actually, there are so many techniques of data analysis making the mind overwhelmed to choose the appropriate ones. Here are some common approaches-Here are some common approaches:

- Statistical Analysis:** Here, you are able to explore measures like mean, median, standard deviation and hypothesis testing to summarize and prepare data. Among the means to investigate causal factors, it reveals these relationships.
- Machine Learning:** Algorithms depend on a priori data to discover behaviors and predictively act. It is for these jobs that the categorization (the task of classifying data points) and regression (the job of prediction of a continuous value) of the data fits well.
- Data Mining:** What's more, it means the exploration of unknown behaviors and occurrences in immense clusters of data. Techniques like association rule learning and clustering cater for identification of latent connections.
- Data Visualization:** Charts, graphs, and dashboards which happen to be tools of visualization of data, make items, trends, and disclosures that would seem to be unclear in raw numbers

3. Explore and Clean the Data

Prior to engaging in any kind of deep analysis, it is vital to grasp the nature of data. EDA takes under analysis the construction of profiles, discovery of missing values, and graphing distributions, in order to figure out what the entire data are about. The data cleaning process allows you to correct inconsistencies, errors and missing values which helps to produce a clear picture based on high quality information.

4. Perform the Analysis

Once all the techniques have been chosen and the data cleaning took place then you can go straight to the data processing itself. Among other techniques, this could encompass performing certain tests, which can be advanced regression or machine learning algorithms, or well-crafted data visualisations.

5. Interpret the Results

You should extract the meaning of the analytics carefully as they are specific to the objectives you have set for yourself. Do not just build the model, show what they signify, make a point by your analysis limitations, and use your starting questions to make the conclusions

6. Communicate Insights

Data analysis is customarily done to advance the decision making. Communicate findings truthfully to all stakeholders such as through means of reports, presentations or interactive charts.

895

♥

↶

🚩

⌂

Make a copy

Create your own with AI

Download PDF

Slide 8

Introducing , Prezi AI. Your new presentation assistantRefineenhance,and tailor your content source relevant imagesyand edit visuals quicker,than ever before.wes
rearsEnabling Technologies for Data Science-I dATA mININGos MOOE)Make a copy, Create your own with AI Download PDF)

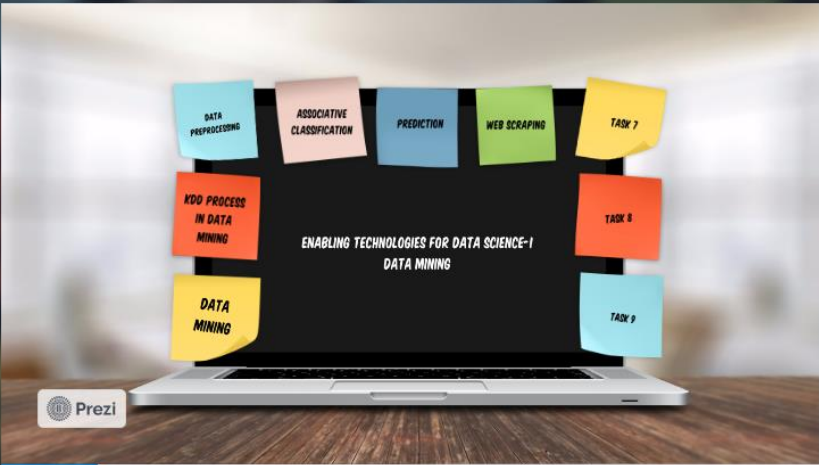
Introducing :+ Prezi AI. Your new presentation assistant. Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker than ever before. [See how →](#)

Prezi

Search presentations...

Log in [Get started](#)

TRENDING SEARCHES [Timeline](#) [Lesson Plan](#) [Great Gatsby](#) [Template](#) [Brainstorm](#)



Enabling Technologies for Data Science-I dATA mINING

895 [❤️](#) [↶](#) [🚩](#) [<>](#)

[Make a copy](#) [Create your own with AI](#) [Download PDF](#)

Slide 9

Introducing , Prezi AI. Your new presentation assistant Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker, than ever before.) Search presentations... KDD Process DD (krmaledon Discovery in Doors Wea process st res the erwarin of valu provauaymie, and potentially vate Ifrnstin rom tao daar Tho KDD proces ea Rateproses ade equres muplo erations ef he abow stops te exact auto knwadge fon the f Foemtion daTheftoving saps waded In KOO process: uw nd i a cee ta aBota tagratonDot harem Is oenad a Natrpenour 3 fram mice soos combina In 2 ca9n Data mining: the core ofsortotarenuh. Dt Ierstr ug te Mr ok, Oot Syn ls nd inowledge discoveryer eerartantctondematn eet.bea BelectonDetssectnn defo the pores ubere dea carte tannin decd mde omtied alo Fr Dh an 86 Nel Dh aon Te Mav bat Chaagrcon mates,a gE TTEct oxhe ros fwmnrsng dit porate fr ese byiom Prezl slemesniee roscoe1: Mop ning amen buco apie taser, Enabling Technologies for Data Science-I dATA mININGos MOOE) Make a copy, Create your own with AI Download PDF)

Introducing :+ Prezi AI. Your new presentation assistant.

Refine, enhance, and tailor your content, source relevant images, and edit visuals quicker than ever before.

See how →



Search presentations...



Log in

Get started

TRENDING SEARCHES

Timeline

Lesson Plan

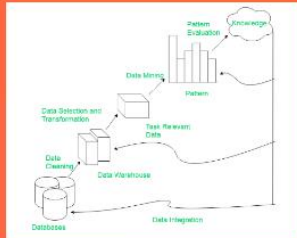
Great Gatsby

Template

Brainstorm

Knowledge Representation

This involves presenting the results in a way that is meaningful and can be used to make decisions.



Note: KDD is an iterative process where evaluation measures can be enhanced, mining can be refined, new data can be integrated and transformed in order to get different and more appropriate results. **Preprocessing of databases** consists of **Data cleaning** and **Data Integration**.



Advantages of KDD

1. **Improves decision-making:** KDD provides valuable insights and knowledge that can help organizations make better decisions.
2. **Increased efficiency:** KDD automates repetitive and time-consuming tasks and makes the data ready for analysis, which saves time and money.
3. **Better customer service:** KDD helps organizations gain a better understanding of their customers' needs and preferences, which can help them provide better customer service.
4. **Fraud detection:** KDD can be used to detect fraudulent activities by identifying patterns and anomalies in the data that may indicate fraud.
5. **Predictive modeling:** KDD can be used to build predictive models that can forecast future trends and patterns.

Disadvantages of KDD

1. **Privacy concerns:** KDD can raise privacy concerns as it involves collecting and analyzing large amounts of data, which can include sensitive information about individuals.
2. **Complexity:** KDD can be a complex process that requires specialized skills and knowledge to implement and interpret the results.
3. **Unintended consequences:** KDD can lead to unintended consequences, such as bias or discrimination, if the data or models are not properly understood or used.
4. **Data Quality:** KDD process heavily depends on the quality of data, if data is not accurate or consistent, the results can be misleading.
5. **High cost:** KDD can be an expensive process, requiring significant investments in hardware, software, and personnel.
6. **Overfitting:** KDD process can lead to overfitting, which is a common problem in machine learning where a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new unseen data.

Enabling Technologies for Data Science-I dATA mINING

895



Make a copy

Create your own with AI

Download PDF