

Subject Incharge: Prof. Sarala Mary Page No.

PARSHWANATH CHARITABLE TRUST'S

## A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering
Data Science



Department of CSE-Data Science | APSIT

Subject : DAV Academic Year: 20 23 20 24 . ROOTS OF TEXT MINING: Text mining is a component of data mining that deals with unstructured lexit data. It involves the use of Natural Planguage Processing CNLP) techniques to extract useful information and insighte from large amount of unstructured text- data. Text mining can be need to extract structured information from une nuctured tent dala such as: (1) Named Entity Recognition (NFR): Identifying and classifying named entities such aspeople, organisations, and locations in tent date. (2) Sentimen Analysis: Identifying and excluding sentiment (eg. positive, negative, neutral) of lint data. Procedure for Analysing Test Mining:. (i) Text summan sation To extract the positial content neplects the whole content automatically. (ii) Text categorisation: To assign a category to the text among categories predefined by users.

PARSHWANATH CHARITABLE TRUST'S



## A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering
Data Science



Semester: VI Subject: DAV Academic Year: 2028-20 24
depending on the several clusters,
depending on their eubetantial relevance.
Icsues in Text Mining:
Numerous issues happen during the text mining proces
* The efficiency and effectiveness of decision-making.
* The uncertain problem can come at an intermediate
stage of text mining. To make the sterd mining process
efficient, different quidelines are developed to
normal the text, in the pre-processing stage.
* Sometimes original message or meaning can be
* Sometimes original message or meaning can be changed due to attention.
* Many algorithms and techniques support Multi- language text, in text mining. And this creates ambiguit
language text, in text mining. And this creates ambiguit
n text mining.
* The use of synonym and antonyms in the document
ent-makes ambiguity in the text mining looks, since int-mining looks take both in a similar setting.  * Hence it is difficult to categorise such kinds of
ent-mining look take both in a similar entire
* Hence it is difficult to categorise such kinds of
int/words.
arge amounts of Data: Pent mining allows organisations  ential-insights from lenge amounts of untrustrual data  Subject Incharge: Prof. Sarala Mary Page No. 2. Department of CSE-Data Science   APSIT
ential-insights from lenge amounts of untructual data
Subject Incharge: Prof. Sarala Mary Page No. 2 Department of CSE-Data Science   APSIT





## A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering
Data Science



Semester: VI Subject: DAV Academic Year: 2093- 20 24 .
It includes social media poste, news articles and customer
Jeedback.
Variety of Applications: Text mining has lot many applications, such as sentimental analysis, named entity recognition and lopic modelling.
such as sentimental analysis, named
Organization can gain insights from une ructured tent
data using the above rematile tool
Improved Decesion Making: Text mining can be used to
Improved Decesion Making: Text mining can be used to extended enrights from unitrustured data and that can be extended enrights from unitrustured data and that can be
used to make data-driven decisions.  Cost-effective: Text mining is a cost-effective way to extract  insights from unitrustured text data  insights from unitrustured text data
Cost-effective: Text mining
Disadvanlages of Text Mining:  Disadvanlages of Text Mining requires advanced skills in natural  Complexity: Text mining requires advanced skills in natural  language processing and machine learning. This makes it  language processing and machine learning extracted
Text mining requires advanced This makes it
completely barressing and machine reaming
language process.
a complex the accuracy of meight
Quality of Data depends on quality of
language processing a complex process.  a complex process. The accuracy of eneights extracted accordingly of textractions depends on qualify of textractions are vary text-mining depends on qualify of textractions are vary. High computational cost is Text mining needs high computational textractions if it is not feasible.  High computational cost is Text mining needs high empulations.  And for small organizations if it is not feasible.
can vary
can vauy. Text mining neede high composible. High computational cost : Text mining neede high computational cost : Text mining neede high computational cost : Text mining is restricted to extracting gesources. And for small organizations is restricted to extracting sesources. In the left and hence cannot
gesources. The fals Text mining is restricted
High conditional for small organizations it restricted to extracting sesources. And for small organizations is restricted to extracting Limited to text date: Text mining is restricted to extracting Limited to text date: Text mining is restricted to extracting the limited to text date and hence cannot ensights from unstructured text date and hence cannot insights from unstructured text date upper.
eniable from unstructure.
Prisights from the data lypes.  be used with other data lypes.  Subject Incharge: Prof. Sarala Mary Page No. 3 Department of CSE-Data Science   APSIT
Subject Incharge: Prof. Sarala Mary Page No. 3 Department of CSE-Data Science   APSIT