## CS 5593

## Data Mining Intro Video

**CTE preparer**: Angie Calton

**Goal/objective:** Provide overview of course and application

**Course fit:** This will be at the start of the course in the “Welcome to the Course” module

**Notes:** Key info highlighted- Some suggestions made about how to conceptualize some of the script.

|  |  |
| --- | --- |
| **Outline/Script** | **Production notes** |
| Welcome to CS5593 Data Mining.  My name’s Le Gruenwald; I’m a professor in the School of Computer Science at the University of Oklahoma. I am the instructor of this course. | headshot |
| Data mining plays a key role in the information economy, culture and learning of the 21st century. There has been enormous data growth in both commercial and scientific databases due to advances in data generation and collection technologies. | Some ideas:  Electric grid (US) <https://unsplash.com/photos/1lfI7wkGWZ4>  Electric grid (NYC)  <https://unsplash.com/photos/_SFJhRPzJHs>  Radar Screen  Shutterstock: ID  [133116230](http://www.shutterstock.com/pic-133116230/stock-vector-commercial-airplane-s-on-board-radar-displaying-weather-information-vector.html?src=LC6tPjQmnea85U8C-qFX3Q-1-53)  Shutterstock: ID [1136056](http://www.shutterstock.com/pic-1136056/stock-photo-secondary-surveillance-radar-situation-screen-display.html?src=LC6tPjQmnea85U8C-qFX3Q-1-58) |
| For example,  Yahoo has Peta Bytes of web data,  Amazon records several million items per day,  and NASA EOSDIS (Earth Observing System Data and Information System) archives over one Peta Bytes of earth science data per year.  These large amounts of data provide us with tremendous opportunities for improving applications by utilizing data mining. | Options:  \*\*\*Perhaps represent ‘peta bytes’. Peta byte – 10005 or 1015 bytes.  1 PB = 1000000000000000B = 1015bytes = 1000  [terabytes](https://en.wikipedia.org/wiki/Terabyte).  \*\*\*Show data (series of 0011010101) streaming over the following images … like the Matrix….  Yahoo! Public domain logo  <https://commons.wikimedia.org/wiki/File:Yahoo_Logo.svg>  Amazon public domain logo  <https://commons.wikimedia.org/wiki/File:Amazon.com-Logo.svg>  Satellite: <https://unsplash.com/photos/VBNb52J8Trk>  Satellite 2  <https://unsplash.com/photos/8Hjx3GNZYeA> |
| So, what can we use data mining for?  Data mining can be used to perform different tasks with data, such as  pattern finding,  association rules structuring,  classification of objects, | Show shapes and colors at random….  RANDOM    PATTERN FINDING – highlight some pattern within the random assortment    ASSOCIATION RULES STRUCTURING (organize by hierarchy)    CLASSIFICATION (sort by color) |
| creation of clusters of similar objects,  and detection of anomalies. | CLUSTERS (sort by size/shape)    ANOMALIES (highlight those oddly shaped pieces and emphasize/twinkle those outliers) |
| From the commercial point of view, Data mining can help provide better customized services and increase profits.  From the scientific point of view, Data mining can facilitate the automated analysis of massive datasets and the formation of hypotheses.  Data mining therefore can provide great opportunities for us to solve many societal problems. | Suggestions to choose from: (do you have similar video clips?)  Stock market data  Shutterstock: ID: [79462060](http://www.shutterstock.com/pic-79462060/stock-photo-colored-ticker-board-on-black.html?src=a6AyIZa1kz-LHcnB7LQWXQ-1-25)  Shutterstock: ID: [68191825](http://www.shutterstock.com/pic-68191825/stock-photo-display-of-stock-market-quotes.html?src=a6AyIZa1kz-LHcnB7LQWXQ-1-49)  Financial data graph  Shutterstock: ID 195875528  Shutterstock ID: [151408784](http://www.shutterstock.com/pic-151408784/stock-photo-stock-market-data-on-screen.html?src=a6AyIZa1kz-LHcnB7LQWXQ-1-27)  App to check stock market data  Shutterstock ID: 281485034  Shutterstock ID: [281485352](http://www.shutterstock.com/pic-281485352/stock-photo-a-city-worker-analysing-stock-market-data-on-a-monitor.html?src=pp-same_model-281485034-a6AyIZa1kz-LHcnB7LQWXQ-6)  Genetics – Sequence 1  Shutterstock ID: [390240103](http://www.shutterstock.com/pic-390240103/stock-photo-genetic-research-and-development-with-science-data.html?src=o4gfN7I1KlKI_IsHyIcNnA-1-41)  Shutterstock ID: [400356196](http://www.shutterstock.com/pic-400356196/stock-photo-genetic-engineering-as-a-science-concept-art.html?src=o4gfN7I1KlKI_IsHyIcNnA-1-86)  Shutterstock ID: [380145898](http://www.shutterstock.com/pic-380145898/stock-photo-stem-cell-research-and-development-as-a-background.html?src=o4gfN7I1KlKI_IsHyIcNnA-1-20)  Shutterstock ID: [374852128](http://www.shutterstock.com/pic-374852128/stock-photo-biotechnology-genetic-research.html?src=pp-same_artist-374848498-o4gfN7I1KlKI_IsHyIcNnA-2)  Shutterstock ID: [395074213](http://www.shutterstock.com/pic-395074213/stock-photo-biotechnology-genetic-research.html?src=o4gfN7I1KlKI_IsHyIcNnA-1-23)  Genetics – Sequence 2  Shutterstock ID: [379679878](http://www.shutterstock.com/pic-379679878/stock-photo-medical-genetics-or-genetic-dna-abstract-image.html?src=o4gfN7I1KlKI_IsHyIcNnA-1-98)  Shutterstock ID: [280849175](http://www.shutterstock.com/pic-280849175/stock-photo-genetic-engineering-as-a-science-concept-art.html?src=o4gfN7I1KlKI_IsHyIcNnA-1-70)  headshot |
| For example, it can help improve healthcare and reduce healthcare cost;  it can help find alternative energy sources;  and it can help reduce hunger and poverty by increasing agriculture production.  As data continue to increase in diversity and volume, the need for data mining is exploding.  An understanding of data mining techniques and how they work is an extremely useful and necessary skill to have in the field of Data Science. | If you have similar video clips, might be better  **Medical technology with patient**  Shutterstock ID: [149858258](http://www.shutterstock.com/pic-149858258/stock-photo-healthcare-medical-and-technology-doctor-showing-something-patient-on-tablet-pc-in-hospital.html?src=yA0DtA7Ja2mglmFqni-0oA-1-44)  Shutterstock ID: [380913907](http://www.shutterstock.com/pic-380913907/stock-photo-male-doctor-medical-students-or-surgeon-using-digital-tablet-and-laptop-during-the-conference.html?src=pp-same_artist-381046621-yA0DtA7Ja2mglmFqni-0oA-2)  Shutterstock ID: [380640088](http://www.shutterstock.com/pic-380640088/stock-photo-male-doctor-medical-students-or-surgeon-using-digital-tablet-and-laptop-during-the-conference.html?src=pp-same_artist-381046621-yA0DtA7Ja2mglmFqni-0oA-7)  Solar farm  Shutterstock ID: [114394150](http://www.shutterstock.com/pic-114394150/stock-photo-solar-farm-in-rural-germany.html?src=q9cMeMUct8cdvrjw7Y5VDg-1-12)  Wind farm  Shutterstock ID: [102426145](http://www.shutterstock.com/pic-102426145/stock-photo-wind-turbine-farm-in-the-field-a-renewable-energy-source.html?src=pp-same_artist-103944314-d5IbJN2g5cPsm_WRApFfwg-3)  Shutterstock ID: [103944314](http://www.shutterstock.com/pic-103944314/stock-photo-wind-turbine-farm-in-the-field-a-renewable-energy-source.html?src=d5IbJN2g5cPsm_WRApFfwg-1-50)  Agricultural science (series)  Shutterstock ID: [321075869](http://www.shutterstock.com/pic-321075869/stock-photo-junior-agricultural-scientists-researching-plants-and-diseases-in-greenhouse-with-parsley.html?src=R4fNNqEfzXpE6blo1-90Tg-1-13)  Shutterstock ID: [373530670](http://www.shutterstock.com/pic-373530670/stock-photo-quality-control-young-female-scientist-stselects-new-breed-of-green-salad-optimized-for.html?src=pp-same_artist-373530718-R4fNNqEfzXpE6blo1-90Tg-1)  Shutterstock ID: [373530718](http://www.shutterstock.com/pic-373530718/stock-photo-quality-control-senior-scientist-or-tech-observes-new-breed-of-cress-sprouts-optimized-for.html?src=R4fNNqEfzXpE6blo1-90Tg-1-9) |
| This course introduces you to the field of Data Mining. You will be introduced to the process, concepts and techniques in data mining. You will gain the background needed in order to apply data mining to real-world problems, and you will use a variety of data mining tools. | headshot |
| Specifically, you will learn what data mining is about, what the knowledge discovery process consists of, what role data mining plays in this process,  and what different data mining tasks you can perform and for what kinds of applications.  You will study how to identify different types of data and what techniques you can use to explore and preprocess data for data mining purposes. | Headshot (you could leave headshot throughout this entire segment) However, if you are worried about headshot being too long or if you need other options…..  Title: **Data Mining**  Bullet 1: **knowledge discovery process**  Bullet 2: **tasks**  Bullet 3: **applications**  Bullet 4: **types of data**  Bullet 5: **techniques for exploring data**  Bullet 6: **techniques for preprocessing data**  \*All 6 points should fit on one slide. You could try to see how this would look as a concept map…  clear screen |
| You will study different techniques to solve the four major data mining tasks:  classification,  clustering,  association analysis, and  anomaly detection. | Refer back to the images of 4 kinds of DM tasks. Add each one to screen, one by one until all four are visible. |
| You will also study the characteristics of Big Data, advanced data mining techniques that deal with Big Data, and Big Data frameworks where those techniques are implemented, ~~such as Hadoop and Spark.~~ | Headshot  \*\*\*make sure to delete audio “…such as Hadoop and Spark”. We don’t want to mention specifics because they may change every year. |
| Please make sure to read the syllabus carefully. It contains important information about the textbooks, grading and other policies.  I hope you will enjoy the course. | Headshot |