

15CSE401

Machine Learning and Data Mining

Lecture-1
15 Jul 2020

Welcome to the course

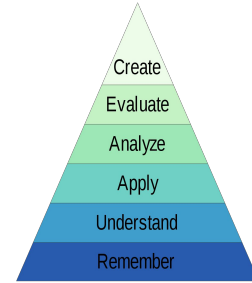
- Course code: 15CSE 401
- Title: Machine Learning and Data Mining
- Semester: 7
- Batch : CSE- D
- Slots:
 - Tuesday : Slot -1 (9-10)
 - Tuesday: Slot - 9 (18-19) (Discussion)
 - Wednesday: Slot -2 (10-11)
 - Saturday: Slot-4 (12- 13)

Course overview

- Quick review of Machine Learning and Data mining concepts
- Advanced topics from ML and Data Mining
- Detailed discussion on statistical and probabilistic concepts
- Applications of concepts in
 - Recommendation systems
 - Text Mining
 - Bioinformatics
 - Web mining
 - ...
- Case study

Course outcome

	Course Outcome	BTL
CO1	Understand the fundamental concepts of Data mining and basic theory underlying Machine learning.	L2
CO2	Understand the types of the data to be mined and apply pre-processing methods	L3
CO3	Apply appropriate classification and clustering techniques for real word applications	L3
CO4	Analyze the performance of various classifiers and clusters techniques	L4
CO5	Apply and evaluate the interesting patterns discovered from association mining	L4



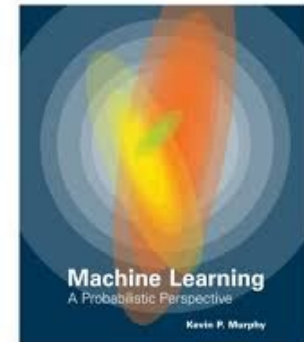
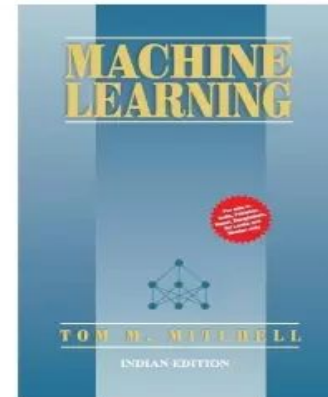
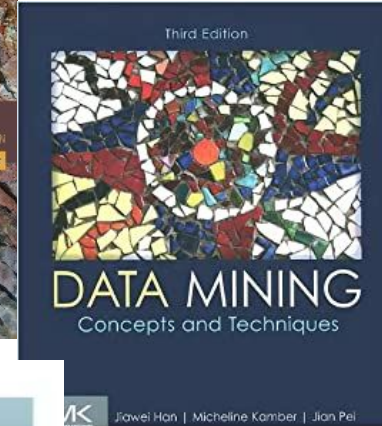
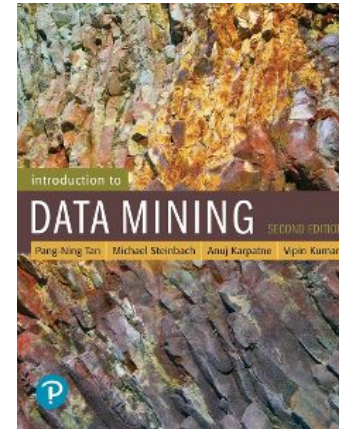
Course overview

Main Topics

- Data mining- Introduction, tasks and challenges
- Measure of similarity and dissimilarity
- Statistical concepts
 - Distributions
 - P-value statistics
- Association Rule mining
- Evaluation of association rule mining
- Bagging and Boosting
- Clustering
 - DBSCAN
 - Fuzzy clustering
 - Hierarchical clustering
 - Cluster evaluation
- Random Forest
- Feature selection and Dimensionality Reduction
- Applications

Text Books

1. Jiawei Han and Micheline Kamber, Jian Pei, “**Data Mining: Concepts and Techniques**”, Third Edition, Elsevier, 2012.
2. Kevin P. Murphey, “**Machine Learning, a probabilistic perspective**”, The MIT Press, 2012.
3. Tom Mitchael, “**Machine Learning**”, McGraw Hill, 1997
4. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, “**Introduction to Data Mining**”, First Edition, Pearson Education, 2006.



Course overview

Course Delivery

- Online classes (MS Teams)
 - Live lectures 20 to 30 minutes
 - Discussion
 - Quiz
- Tutorial/Discussion hour (MS Teams)
- Weekly Quizzes (AMPLE / AUMS)
- Assignments (Problems and programming)
- Case study (Group)
- Course Repository - AMPLE



Course overview

Evaluation Pattern

- | | |
|-----------------------|------------|
| ● Quiz (Teams/ Ample) | 20% |
| ● Assignments | 30% |
| ● Case study | 20% |
| ● End Exam (online) | 15% |
| ● Viva | 15% |

Course overview

General comments

- **Keep a separate course note book**
- **Active participation** - Chats and discussion forums
- **Be on time** in all evaluations
- **Contact info:**
 - m_manu@cb.amrita.edu
 - 7736735850

Data Mining

— Introduction —

Examples



- ? Where should detergents be placed in the Store to maximize their sales?
- ? Are window cleaning products purchased when detergents and orange juice are bought together?
- ? Is soda typically purchased with bananas? Does the brand of soda make a difference?
- ? How are the demographics of the neighborhood affecting what customers are buying?

Examples

Customers who viewed this item also viewed

Page 1 of 6



Sony WI-C200 Wireless In-Ear Headphones with 15 Hours Battery Life, Quick Charge, Magnetic...
★★★★☆ 8,339
₹ 1,890.00 ✓prime



boAt Rockerz 255 Sports Wireless Headset with Super Extra Bass, IPX5 Water & Sweat...
★★★★☆ 1,10,937
₹ 1,499.00 ✓prime



JBL T205BT Pure Bass Wireless Metal Earbud Headphones with Mic (Black)
★★★★☆ 10,940
₹ 1,499.00 ✓prime



Realme Buds Wireless
★★★★☆ 25,475
₹ 1,789.00 ✓prime



Noise Tune Active Bluetooth Wireless Headset with Upto 10 Hour Playtime, IPX5...
★★★★☆ 224
₹ 1,499.00



boAt Rockerz 239 Wireless Headset with ASAP Fast Charge, Immersive Audio, Bluetooth V5.0,...
★★★★☆ 307
₹ 1,299.00 ✓prime



(Renewed) Infinity Glide 120 Metal In-Ear Wireless Flexband Earphones, with Bluetooth 5.0 and IPX5...
★★★★☆ 31
₹ 1,199.00 ✓prime



Customers who bought this item also bought

Page 1 of 7



Clavier Hard Carrying Case
Portable Protection



Infinity Glide 100 Wireless In-Ear Dual EQ Deep Bass



Infinity Zip 100 In-Ear Immersive Bass Tangle



Tizum Z13 Earphone Carrying Case for



TIZUM Earphone Carrying Case - Multi Purpose



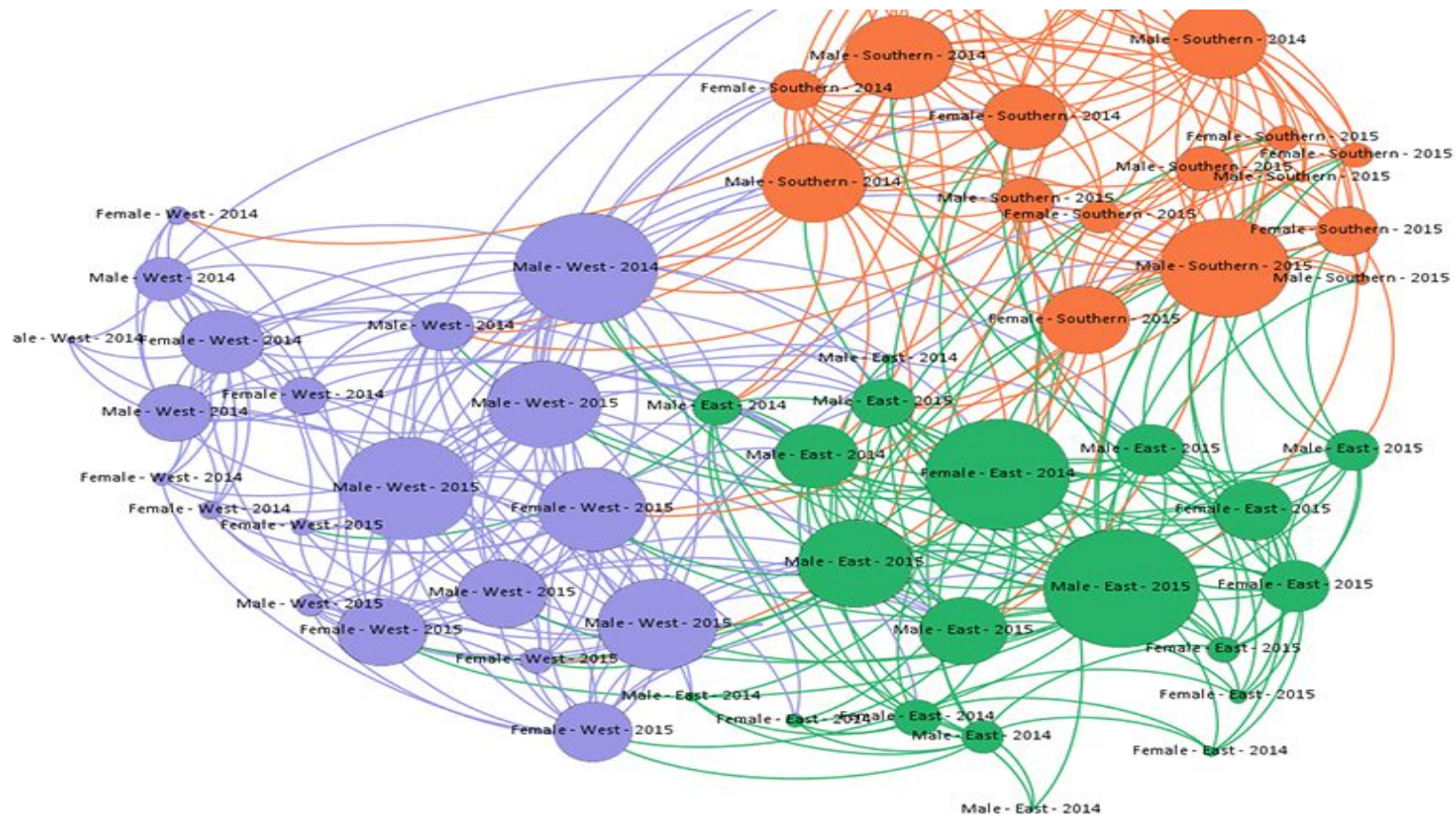
Infinity(JBL) Glide 510 On-Ear Wireless Headphone



JBL C100SI In-Ear Deep Bass Headphones with Mic



Examples



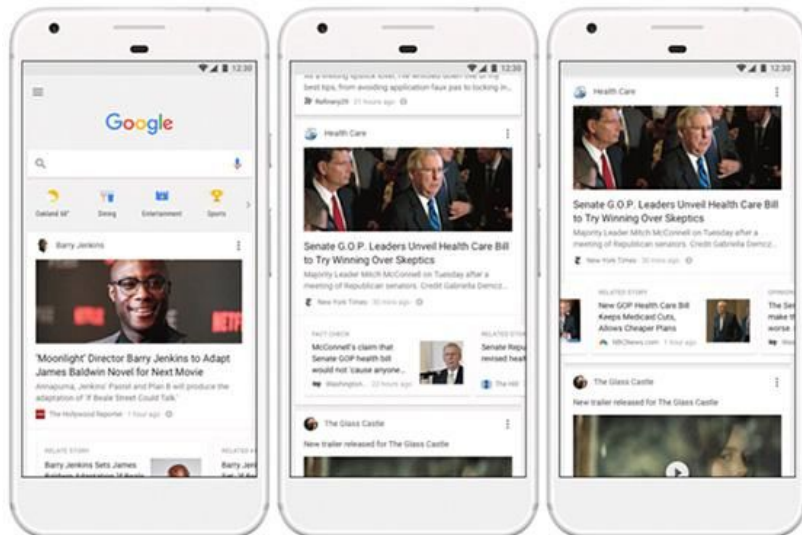
Examples

 You have a friend suggestion. What's this?

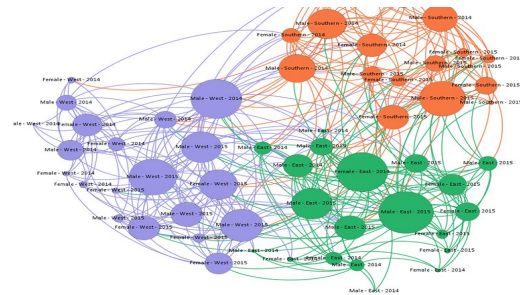
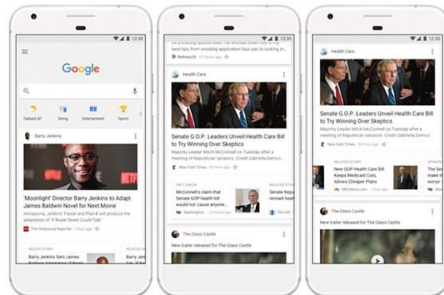
Friend Suggestions

Your friend suggestions are generated when one of your friends select you as someone who knows someone else on Facebook. If you add your suggested friends as friends, a normal friend request will be sent. If you do not, no one will be notified that you ignored a suggestion.

Okay



What is common in all??



Customers who viewed this item also viewed



Customers who bought this item also bought



Data Mining

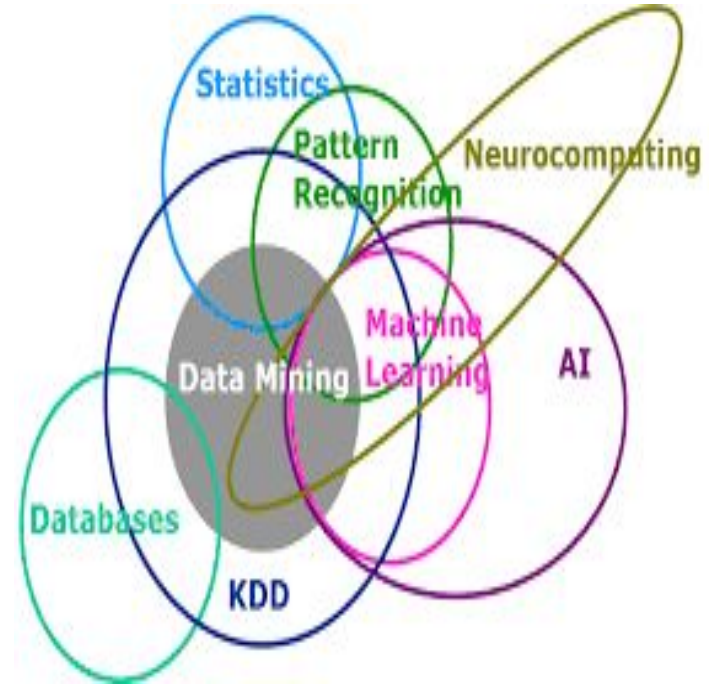
- We are living in '**BIG DATA**' age
 - **Data is wealth**
- Explorative growth of data from tera to peta bytes
- Major sources
 - Business - Web, e-commerce, transaction databases
 - Science - Remote sensors, Biology, various simulations
 - News, images, chats, Tweets, social media,...
 - Everyone contributing
- We are drowning in data, but starving for knowledge!- Eric D Brown

Data Mining

- Data → Information → Knowledge → Wisdom
- Manual analysis of big data is not practical
- **Data mining—Automated analysis of massive data sets**
- **Data mining is a process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems-** wikipedia
- Data mining turns a large collection of **data into knowledge**

Data Mining

- **Data mining** is performed by humans on certain data sets with the aim to find out interesting patterns between the items in a data set [softwaretestinghelp.com (blog)]
- Data mining uses techniques developed by **machine learning** for predicting the outcome
- It is a Multi-disciplinary area



Data Mining

- Mining frequent patterns
 - Association mining
 - Correlation analysis
- Classification
 - Predictive analysis
 - Regression
- Clustering
 - Descriptive analysis
 - Community detection
- Outlier analysis
 - Anomaly detection

Next

- Detailed discussion on data mining tasks
- Challenges in Data mining
- Examples

Reading Assignment

- Chapter -1, Jiawei Han and Micheline Kamber, Jian Pei, **“Data Mining: Concepts and Techniques”**, Third Edition.

Thank you