

Data Mining

Prepared by Dr. Amjad Qtaish

DS601
Data Mining

Orientation Lecture



Data Mining (DS601)



Dr. Amjad Qtaish

Faculty of Informatics
Midocean University



عرض توجيہي Orientation show

Who am I?



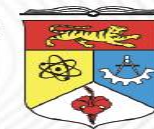
Dr. Amjad Faisal Ali Qtaish
Doctor of Philosophy (PhD) Degree
of Computer Science / Data Science



EDUCATION

■ **PhD Degree of Computer Science/ Data Science**

The National University of Malaysia
2016
Bangi, Selangor, Malaysia



UNIVERSITI
KEBANGSAAN
MALAYSIA
*National University
of Malaysia*

■ **Master Degree of Computer Science**

University Putra Malaysia
2013
Serdang, Selangor, Malaysia



UPM
UNIVERSITI PUTRA MALAYSIA
BERSAMA SAMA MELAKSANAKAN TRANSFORMASI

■ **Bachelor Degree of Computer Science**

Zarqa Private University
2002
Zarqa, Jordan



Who am I?



PHD THESIS

■ XAncestor

A Mapping Approach for Storing and Querying XML Documents in Relational Database using Path-Based Technique



ACADEMIC AWARDS

■ Certificate of Achievement

The National University of Malaysia 2016

■ Outstanding Researcher Award

The National University of Malaysia 2016



UNIVERSITI
KEBANGSAAN
MALAYSIA
*National University
of Malaysia*



UNIVERSITI
KEBANGSAAN
MALAYSIA
*National University
of Malaysia*

Who am I?



SUMMARY OF QUALIFICATIONS

- Bsc, Master, and PhD Degrees in Computer Science (Programming, Web Technology, Database, Big data, Distributed Systems and Networking) / **Data Science**
- Lecturer in Computer Department, Al-Kharj College of Technology, Saudi Arabia 2004 - 2013
- Head of Computer Department, Al-Kharj College of Technology, Saudi Arabia 2006- 2008
- Teaching and Research Assistant in FTSM Faculty, The National University of Malaysia, Bangi, Malaysia 2014 - 2016
- Publishing four High Quality of Journal Papers that indexed by Scopus and ISI.
- Excellent Programming using Oracle Developer 9i & Others Languages
- Administering LAN & WAN networks in Windows 2003 , 2000 platforms
- Administering Different Types of RDBMS such as Oracle, SQL Server, MySQL, SQLite
- Excellent Work in Management Depend on Previous Experience
- Certificate for IT Essentials: PC Hardware and Software
- Certificate for CCNA Exploration : Network Fundamentals
- Certificate for CCNA Exploration : Routing Protocols and Concepts
- Certificate of Participation as Presenter in International Conference on Electrical Engineering and Informatics (ICEEI 2015), Bali, Indonesia
- Certificates of Participations in two Scientific Workshops in FTSM, the National University of Malaysia
- Participant in two Scientific Seminars in FTSM, the National University of Malaysia
- Award - Certificate of Achievement from the National University of Malaysia 2016
- Award -Outstanding Researcher from the National University of Malaysia 2016
- Programming Many Software Projects During my Study and my Work
- Presenting Many Training Courses for Lecturer in Many Courses in Computer Science During my Work such as SQL, PL/SQL, Java, VB.Net, ASP, Java, and Typing Skills

My Contact Information



- ❖ Office hours
 - Discussion Time (Wednesday at 6:30 PM) every week
- ❖ More about me:



<https://scholar.google.com/citations?user=JejNDrcAAAAJ&hl=en>



<https://www.scopus.com/authid/detail.uri?authorId=56720387400>



<https://orcid.org/0000-0001-5210-7086>

The course that will be taught



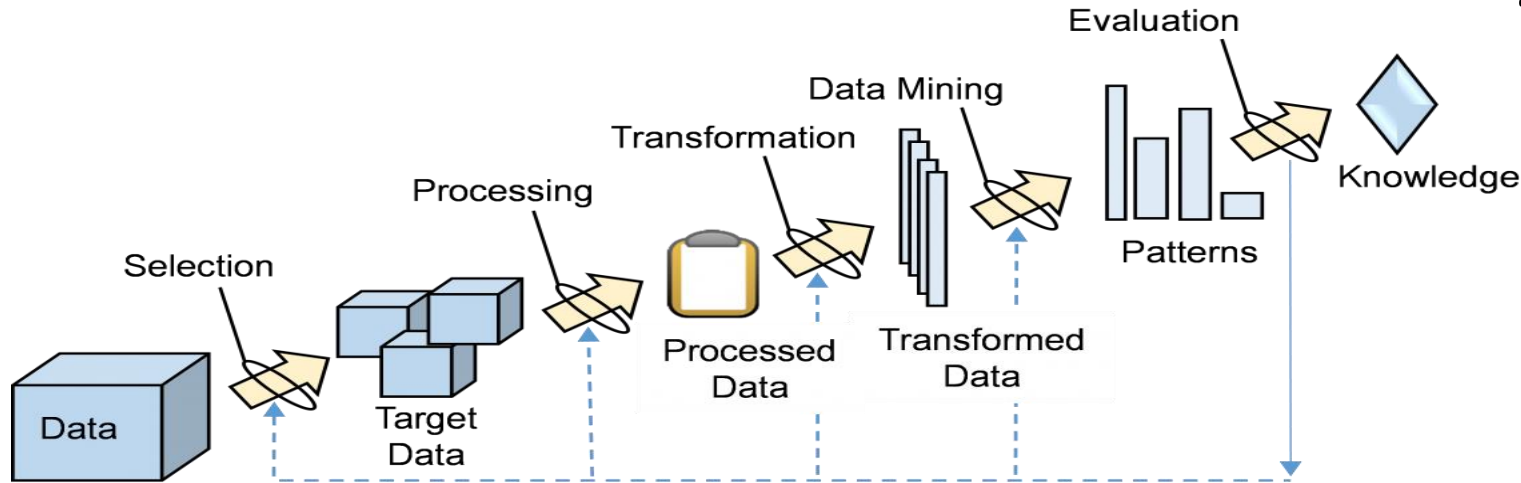
Data Mining



التنقيب عن البيانات وأهميته Importance of Data Mining



يشير معنى التنقيب عن البيانات Data mining الى عملية استخراج الأنماط أو الأفكار المفيدة من كميات كبيرة من البيانات. ويتضمن استخدام تقنيات وخوارزميات مختلفة لاكتشاف الأنماط والاتجاهات المخفية داخل البيانات. بهدف استخراج معلومات قيمة يمكن استخدامها لاتخاذ القرار، حيث يساعد المؤسسات على الكشف عن رؤى قيمة، وتحسين عمليات صنع القرار، وتحديد أنماط سلوك العملاء، واكتشاف الأنشطة الاحتيالية، وتحسين العمليات، وتحسين أداء الأعمال بشكل عام.



Primary Goal of This Course



يركز التنقيب عن البيانات على إيجاد أنماط جديدة ومثيرة للاهتمام وربما مفيدة في البيانات. ويهدف إلى تحويل كمية كبيرة من البيانات إلى معرفة. أصبح استخراج البيانات مجالاً مهماً جداً في الصناعة وكذلك في الأوساط الأكاديمية. يغطي هذا المقرر معظم تقنيات التنقيب عن البيانات المهمة وتوفر معرفة أساسية حول كيفية إجراء مشروع التنقيب عن البيانات. بعد ذلك، ستم مناقشة التقنيات الأساسية لكيفية المعالجة المسبقة لمجموعة بيانات لمهمة استخراج البيانات، وتحديد ماهية اكتشاف المعرفة واستخراج البيانات، ومهام استخراج البيانات مثل التصنيف والتجميع وتحليل الارتباطات بالتفصيل.

Data mining centers on finding novel, interesting, and potentially useful patterns in data. It aims at transforming a large amount of data into a well of knowledge. Data mining has become a very important field in industry as well as academia. The course covers most of the important data mining techniques and provides background knowledge on how to conduct a data mining project. After, basic techniques of how to preprocess a data set for a data mining task, defining what knowledge discovery and data mining is, data mining tasks such classification, clustering, and association analysis will be discussed in detail.

What This Course is About



➤ The course will help students to gauge understanding in :

- **Data cleaning and preparation:** Students will learn how to identify and address errors, inconsistencies, and missing values in data sets to ensure their quality for analysis.
- **Data analysis techniques:** Courses will cover various methods for analyzing data, such as statistical analysis, machine learning algorithms, and data visualization techniques.
- **Pattern discovery:** A core aspect of data mining is finding hidden patterns and relationships within data. Students will learn algorithms for tasks like classification, clustering, association rule learning, and anomaly detection.
- **Communication and storytelling:** Effectively presenting insights from data is crucial. Courses may cover techniques for data visualization and communication to translate complex findings into clear and actionable results.



Units of Course Contents

Module	Course Unit	No.of. Weeks	Contact Hours
1	Lecture 1: Introduction to Data Mining	1 st	3
2	Lecture 2: Data Preprocessing	2 nd	3
3	Lecture 3: Data Visualization	3 rd	3
4	Lecture 4: ETL (Extract Transform Load)	4 th	3
5	Lecture 5: DQM (Data Quality Management)	5 th	3
6	Lecture 6: Mining Frequent Patterns, Associations, and Correlation	6 th	3
7	Lecture 7: Classification and Prediction	7 th	3
8	Lecture 8: Classification and Prediction using Orange Tool	8 th	3
9	Lecture 9: Dimensional Modeling	9 th	3
10	Lecture 10: Cluster Analysis	10 th	3



Recorded Lectures
المحاضرات المسجلة

Live Sessions
جلسات وحلقات النقاش

Virtual Labs
المعامل الافتراضية

References
المراجع المعتمدة للمقرر الدراسي

منهجية التدريس



يتم احتساب الغياب من خلال حل الطالب للاختبار القصير المصاحب للمحاضرة الأسبوعية المسجلة، وفي حال تغيب الطالب عن ٤ اختبارات قصيرة يعتبر الطالب في هذه الحالة حرمان.

آخر موعد للاعتذار عن أي اختبار قصير يكون قبل إغلاق موعد الاختبار مع إرفاق الدلائل الثبوتية لأي أعذار قهرية أو مرضية عبر نموذج الأعذار عن اختبار قصير عبر منصة (SIS).

آخر موعد للاعتذار عن الاختبار النهائي يكون قبل موعد عقد الاختبار مع إرفاق الدلائل الثبوتية لأي عذر قهري أو طبي عبر نموذج الأعذار عن اختبار نهائي عبر منصة (SIS)، ولن يتم احتساب أي أعذار بعد نهاية الموعد، ويحق لإدارة الجامعة مراعاة بعض الظروف القهرية.

www.midocean.ae



الأعمال وتوزيع درجات الفصل الدراسي الثالث 2023 - لمرحلي البكالوريوس والماجستير

توزيع الدرجات - 100%

حضور المحاضرة - حل الأسئلة المصاحبة

8 درجات لكل اختبار قصير -
بمجموع 80 %

الحضور اختياري

20%

الأعمال

محاضرة مسجلة أسبوعيًا -
السبت من كل أسبوع

اختبار قصير مصاحب للمحاضرة
بشكل أسبوعي

حلقة نقاش افتراضية أسبوعية
لمناقشة محتوى المحاضرة المسجلة

الاختبار النهائي



➤ Assessment Method:

▶ Short Exam

- ▶ **80% marks**
 - ▶ Count of short Exam 10 (Weekly)
 - ▶ 8 Marks for each of short Exam.
- ▶ **20% marks**
 - ▶ Final Examination





➤ **After completion of this course, the students should be able to:**

- Identify and remove missing values, noise and outliers in presented data.
- Characterize and rank the association rules using support, confidence, and lift.
- Understand how classification methods can be used to perform predictive data mining.
- Validate a prediction model using split and cross validation methods.
- Compare and contrast classifiers using various evaluation measures.
- Differentiate between hierarchical and partitioning clustering methods.



➤ **Credit Hours:**

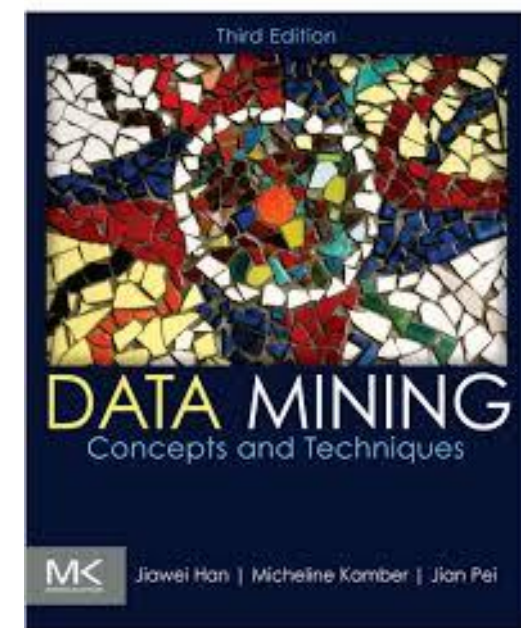
3 credit hours.

➤ **Required Textbook:**

Jiawei Han, Jian Pei, Hanghang Tong, (2012). Data Mining Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems. Third Edition.

➤ **Essential References:**

Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition.





Thank You

