



שם הקורס בעברית: מדעי הנתונים ובינה עסקית

שם הקורס באנגלית: Data Science and Business Intelligence

מספר הקורס: 372-1-3105

מבנה הקורס: הרצאה: 3, תרגיל 1, מעבדה 1, סה"כ: 5 ש"ש

נקודות זכות: 4.0

**מרצה: פרופ' מרק לסט
מתרגלים: גב' ניבה חזון, מר גיל אברהמי
סמסטר ב', תשע"ט**

תאור הקורס:

הקורס נפתח עם מבוא לתחום של מדעי הנתונים (Data Science). לאחר מכן נמשיך עם סקירה של טכניקות להכנת נתונים והנדסת נתונים. כמו כן, נספק את הרקע הדרוש בתורת האינפורמציה. בחלק המרכזי של קורס, נכסה את השיטות הנפוצות ביותר לכריית נתונים ולמידה חישובית כולל עצי החלטה, רשתות אינפ-עמומות, רשתות נוירונים מלאכותיות, למידת בייס, למידה מבוססת תצפיות, מכונות של ווקטורי תמיכה (SVM), חוקי קשר, וניתוח אשכולות. נסיים את הקורס עם מושגי היסוד בבינה עסקית (Business Intelligence) כגון מחסני נתונים ועיבוד אנליטי מקוון (OLAP). בשיעורי תרגול, התלמידים יפתרו בעיות מעשיות באמצעות טכניקות ואלגוריתמים שלמדו בכיתה. בשיעורי המעבדה, התלמידים יתנסו בשימוש בכלים הפופולריים ביותר לכריית מידע: R ו-Python.

מטרות הקורס:

ללמוד ולתרגל את השיטות המובילות בתחומים של מדעי הנתונים ובינה עסקית כולל יישומן לפתרון בעיות בעולם האמיתי.

תנאי קדם:

37211021 מבוא להסתברות וסטטיסטיקה

37212021 אמידה ובדיקת השערות

37213305 בסיסי נתונים



דרישות הקורס והרכב הציון:

- הקורס מתבסס על הרצאות בכיתה, שיעורי תרגול, שיעורי מעבדה וחומר הקריאה. **המצגות המפורסמות באתר הקורס אינן מכסות בהכרח את כל החומר הנלמד בכיתה.**
- בחנים מקוונים: כ-5 פעמים במהלך הסמסטר, יפורסם באתר הקורס בוחן מקוון המכסה את החומר שנלמד בשיעורים הקודמים. התלמידים חייבים לבצע את הבחנים המקוונים עד מועדי ההגשה שיפורסמו באתר. זו אחריותו האישית של כל תלמיד לבדוק באתר הקורס על בסיס קבוע את ההודעות על פרסום בחנים חדשים. משקל כולל של כל הבחנים: כ-10%.
- עבודות בית: יפורסמו כחמש עבודות בית במשקל כולל של כ-20% מהציון הסופי. משקל כל עבודה יפורסם באתר קורס. את עבודות הבית יש להגיש ביחידים או בזוגות במועדים שייקבעו. הם יכללו מטלות חישוביות ותכנותיות.
- מפגשי מעבדה: כל מפגש מעבדה יהיה כרוך בגשת דו"ח קצר בצורת קוד או פלט תוכנה. דרישות מפורטות יפורסמו באתר הקורס. המשקל הכולל של משימות אלה – כ-10%.
- בחינה מסכמת: בסוף הסמסטר, תתקיים בחינה סופית. משקל: כ-60%.
- כל מטלה שלא תוגש בזמן ללא סיבה מוצדקת (כגון מחלה או מילואים) תקבל ציון של אפס.
- השתתפות בכל מפגשי המעבדה והצלחה בבחינה הסופית נדרשים למעבר הקורס.

רשימת קריאה (ביבליוגרפיה):

קריאות חובה

- T1 Han, J., Kamber, M., and Pei, J. [*Data Mining: Concepts and Techniques*](#), 3rd Edition, Morgan Kaufmann, 2011.
- T2 Kelleher, J. D., Mac Namee, B., & D'Arcy, A., [*Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*](#). MIT Press, 2015.
- T3 Hastie, T., Tibshirani, R., and Friedman, J., [*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*](#), Second Edition, Springer Verlag, 2009.

קריאות רשות

- R1 Mitchell, T.M., [*Machine Learning*](#), McGraw-Hill, 1997.
- R2 Maimon, O. and Last, M., [*Knowledge Discovery and Data Mining – The Info-Fuzzy Network \(IFN\) Methodology*](#), Kluwer Academic Publishers, Massive Computing Series, 2000.
- R3 E. Sperley, *Enterprise Data Warehouse: Planning, Building, and Implementation*, Volume 1, Prentice-Hall, 1999.
- R4 Pyle, D., *Data Preparation for Data Mining*, Morgan Kaufmann, San Francisco, CA, 1999.



תכנית המפגשים:

שיעור מס'	נושא השיעור	קריאה נדרשת
1	מבוא למדעי הנתונים וגילוי ידע בבסיסי נתונים	T1, Ch. 1 T2, Ch. 1
2	הכנת הנתונים והנדסת נתונים	T1, Ch. 3 T2, Ch. 3
3	מבוא לתורת האינפורמציה	T2, Ch. 4 R2, Sec. 1.7 + Appendix A
בוחן מקוון מס' 1		
4-5	למידה באמצעות עצי החלטה	T1, Sec. 8.2
6	רשתות אינפו-עמומות	R2, Ch. 3
בוחן מקוון מס' 2		
7	רשתות נוירונים מלאכותיות	T1, Sec. 9.2 T3, Ch. 11
8	שיטות למידה בייסאניות	T1, Sec. 8.3, 9.1
9	למידה מבוססת תצפיות ומכונת וקטורים תומכים (SVM)	T1, Sec. 9.3, 9.5 R1, Ch. 8 T3, Ch. 12
בוחן מקוון מס' 3		
10	גילוי של חוקי קשר	T1, Ch. 6
11	ניתוח אשכולות	T1, Ch. 10 T3, Ch. 14.3
בוחן מקוון מס' 4		
12	מבוא לבינה עסקית (BI), מתודולוגיה של מחסני נתונים	T1, Sec. 4.1 T2, Ch. 2
13	ניתוח אנליטי מקוון (OLAP), Big Data, Predictive Data Analytics	T1, Sec. 4.2 T2, Ch. 11
בוחן מקוון מס' 5		