



The Hebrew University of Jerusalem

Syllabus

Big Data and Business Intelligence for Political Science - 56146

Last update 08-03-2023

HU Credits: 4

Degree/Cycle: 1st degree (Bachelor)

Responsible Department: Political Science

Academic year: 0

Semester: 2nd Semester

Teaching Languages: Hebrew

Campus: Mt. Scopus

Course/Module Coordinator: Alon Peled

Coordinator Email: peled.alon@gmail.com

Coordinator Office Hours: Monday 1300-1430

Teaching Staff:

Prof Alon Peled

Course/Module description:

Can we analyze video movies and photos to predict if the Prime Minister or President is lying to the nation? How can we analyze “clues” buried inside the parliamentary speeches of new Knesset members to predict if they will become excellent legislators? The Big Data and AI (Artificial Intelligence) revolution empowers us to answer such questions by streaming huge and diverse quantities of textual, audio, and visual data into ‘data lakes’, integrating the data inside these lakes, and, then, developing and operationalizing machine learning and deep learning models to extract new insights. The evolution of commercial clouds such as GCP (Google cloud), Azure (Microsoft cloud), AWS (Amazon cloud), and Watson (IBM cloud) provides us with a cheap, fast and simple opportunity to develop infrastructures and user interfaces to address questions we did not dare to ask yesterday!

Course/Module aims:

The course’s objective is to endow the students with an understanding of basic concepts in the domains of big data and artificial intelligence. To examine the background, reasons, and history through which these concepts have evolved. To understand the challenges, difficulties, and tradeoff relationships of using software tools that implement these concepts. To experience first hand and for the first time how to work creatively with huge quantities of data in a commercial cloud. All this in order to convince the student to continue working with big data and AI tools after the completion of the course!

Learning outcomes - On successful completion of this module, students should be able to:

At the end of the course, the students will be empowered to do the following:
Define basic concepts such as big data, machine learning, and visual display of quantitative data.

Mine data by writing code in a commercial cloud.

To create and operate Auto-ML models to analyze text, photos, and videos.

Find answers to various programming questions on their own by using tools such as chatgpt and stackoverflow..

Self-learn various big data concepts by using online tutorials.

Acquire initial understanding of advanced topics such as deep learning.

Attendance requirements(%):

90%

Teaching arrangement and method of instruction: Monday - 13:00 - 14:45

Thursday - 08:30 - 10:15

Course/Module Content:

פגישה 1 (שני 13 מרץ):

הקורס: יעדים ותכנית ההדרכה. היכרות. חלוקה לקבוצות (לא יותר מארבעה תלמידים בצוות!).
קריאה: אין

פגישה 2 (חמישי 16 מרץ):

ברוכים הבאים לעולם הנפלא של ביג דאטה ובינה מלאכותית!
(בחירת פרוייקטים צוותיים)

קריאה: 2013 ,Data Big ,Schönberger-Mayer

עיון:

<https://blog.gdeltproject.org/>

פגישה 3 (שני 20 מרץ)

“עשרת הדיברות” לצעדים ראשונים בעולם הנתונים

קריאה: 2016 Destruction Math of Weapons .Neil'O

עיון:

<https://teachablemachine.withgoogle.com/>

פגישה 4 (חמישי 23 מרץ):

היכרות עם כלים מרכזיים בקורס והתקנות

Notepad++, chatgpt, Stackoverflow, Google Cloud Platform

קישורים

<https://openai.com/blog/chatgpt>

<https://notepad-plus-plus.org/downloads/>

<https://stackoverflow.com/>

פגישה 5 (שני 27 מרץ):

מושגי יסוד בעולם פלטפורמת הענן (platform cloud)

פגישה 6 (חמישי 30 מרץ):

פגישה ראשונה עם מאגר ביג דאטה - ג'דלט (Gdelt)

קריאה: 2013 GDELT ,Schrodte & Leetaru

חופשת פסח - 2 - 15 אפריל

פגישה 7 (שני 17 אפריל):

פתיחת חשבונות ענן אישיים בפלטפורמת הענן של גוגל והיכרות ראשונית עם כלי הענן של גוגל

פגישה 8 (חמישי 20 אפריל):
Google bigquery "Hello World" *SELECT...FROM... WHERE...ORDER BY ...LIMIT*
2004 *Basics Database Relational*, Sato & King :קריאה

חופשת יום הזיכרון ויום העצמאות - 24 - 26 אפריל

פגישה 9 (חמישי 27 אפריל):
Google bigquery *GROUP BY ...ORDER BY*

פגישה 10 (שני 1 מאי):
Google bigquery *Working with Arrays (UNNEST... ARRAY_AGG)*

פגישה 11 (חמישי 4 מאי):
Google bigquery *JOINS 1*

פגישה 12 (שני 8 מאי):
Google bigquery *JOINS 2*

פגישה 13 (חמישי 11 מאי):
Google bigquery - *Types and Type-Functions*

פגישה 14 (שני 15 מאי):
Google bigquery - *Advanced Functions*

פגישה 15 (חמישי 18 מאי):
Google bigquery - *Analytical Functions*

פגישה 16 (שני 22 מאי):
Google bigquery - *Creating a Data Pipeline*

חופשת שבועות - 25 - 26 מאי

פגישה 17 (שני 29 מאי):
Google bigquery - *Scheduling the Data Pipeline With Partitioning*

פגישה 18 (חמישי 1 יוני):
Google Data Looker - *Creating a Simple Dashboard*
2001 *Excellence Graphical*, Tufte :קריאה

פגישה 19 (שני 5 יוני):
Google Data Looker - *Displaying Timeline, Trend, and Geography*

פגישה 20 (חמישי 8 יוני):
Google Cloud Videointelligence API - *Analyzing photo*

קריאה: 2018 Background ,Algorithms Learning Machine Master ,Brownlee
עיון:

<https://machinelearningmastery.com/>

<https://www.kaggle.com/learn/intro-to-machine-learning>

<https://carto.com/help/tutorials/getting-started-with-carto-builder/>

פגישה 21 (שני 12 יוני):

Google Cloud Videointelligence API - Analyzing video

פגישה 22 (חמישי 15 יוני):

Google Cloud Auto-ML - Creating a Model to Analyze Photos

פגישה 23 (שני 19 יוני): עבודה מורה-צוותים על הפרוייקטים

פגישה 24 (חמישי 22 יוני): עבודה מורה-צוותים על הפרוייקטים

פגישה 25 (שני 26 יוני): מצגות פרוייקטים א'

פגישה 26 (חמישי 29 יוני): מצגות פרוייקטים ב'

Required Reading:

Booch, Grady. *Object-oriented analysis and design with applications*. Redwood City, CA: The Benjamin/Cummings Publishing Company, Inc. 2007.

Brownlee, Jason. *Master Machine Learning Algorithms - Discover How They Work and Implement Them From Scratch*. 2018. Available at:

<https://machinelearningmastery.com/master-machine-learning-algorithms>.

King, Gary., Keohane, Robert. O., & Verba, Sidney. *Designing social inquiry: Scientific inference in qualitative research*, Princeton: Princeton University Press, 1994.

King, Ro and Sato, Todd. "Relational Database Basics" *Information Management* (2004). Available at <https://www.information-management.com/news/relational-database-basics>.

Leetaru, Kalev and Schrodtt, Philip A. "GDELT: Global Data on Events, Location and Tone, 1979-2012. 2013. Available at <http://data.gdeltproject.org/documentation/ISA.2013.GDELT.pdf>.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray, 2013.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers, 2016. 272p.

Tufte, Edward R., 1942-. *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press, 2001.

Additional Reading Material:

Brownlee, Jason. *Master Machine Learning Algorithms - Discover How They Work and Implement Them From Scratch*. 2018. Available at:
<https://machinelearningmastery.com/master-machine-learning-algorithms>.

King, Ro and Sato, Todd. "Relational Database Basics" *Information Management* (2004). Available at <https://www.information-management.com/news/relational-database-basics>.

Leetaru, Kalev and Schrod, Philip A. "GDELT: Global Data on Events, Location and Tone, 1979-2012. 2013. Available at
<http://data.gdeltproject.org/documentation/ISA.2013.GDELT.pdf>.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray, 2013.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers, 2016. 272p.

Tufte, Edward R., 1942-. *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press, 2001.

Course/Module evaluation:

End of year written/oral examination 0 %

Presentation 0 %

Participation in Tutorials 10 %

Project work 0 %

Assignments 30 %

Reports 0 %

Research project 60 %

Quizzes 0 %

Other 0 %

Additional information:

הערות חשובות:

תלמיד שייעדר מיותר משלוש פגישות מעבדה ללא הסבר מספק לא יקבל ציון עבור הקורס

כל תלמיד/ה יתבקש להגיע לפגישה אישית עם המרצה במהלך הסמסטר.
כל תלמיד/ה נדרש/ת להגיש שני תרגילים. כל תרגיל יוגש למורה שבוע לאחר שהתלמיד/ה מקבל/ת התרגיל.
התלמידים יתבקשו להתקין תוכנות חינוכיות שונות על המחשבים האישיים שלהם. את ההתקנות נבצע ביחד במעבדות הראשונות בקורס.
התלמידים ידרשו לפתוח חשבון אישי בשירותי ענן (למשל: Platform Cloud Google-GCP) תוך שימוש בפרטים אישיים, חשבון ג'מייל לא אוניברסיטאי, וכרטיס אשראי אישי. השימוש במקורות GCP חינם אך ורק אם התלמידים עוקבים בקפדנות אחר הוראות המורה. כל שימוש במקורות GCP שלא על פי הנחיות המורה עשוי לעלות כסף ויהיה על חשבון התלמיד! את חשבון פלטפורמת הענן של גוגל אתם חייבים לפתוח באמצעות חשבון ג'מייל פרטי שלכם (לא תוכלו להשתמש בחשבון הג'מייל של האוניברסיטה העברית למטרה זו). את החשבונות נפתח ביחד במעבדה הראשונה בקורס (בבקשה לא לפתוח חשבונות אלו לפני כן כיוון שהקרדיט חינם לשימוש בהם מוגבל לשלושה חודשים!). אפשר, כמובן, לפתוח חשבון ג'מייל נפרד אך ורק למטרת השימוש ב GCP אולם גם אז תדרשו למסור לגוגל פרטי כרטיס אשראי אישי כדי לפתוח את החשבון. תלמיד או תלמידה אשר מסיבה כלשהי לא יכול לעמוד בתנאי סף זה לא יוכל/תוכל להשתתף בקורס.