

תאריך עדכון: 9/6/2021

## שם ומספר הקורס:

**עיבוד, ניתוח והצגת נתונים בכימיה 84642-01**

### שם המרצה: דן מאיור

סוג הקורס: הרצאה + תרגיל

שנת לימודים: תשפ"ב    סמסטר: א    היקף שעות: 2+2

אתר הקורס באינטרנט: מודל

שעות הקורס: יום ד' 14:00-16:00

### א. מטרת הקורס ותוצרי למידה (מטרות על / מטרות ספציפיות):

מטרת הקורס –

להעניק יכולת מתקדמת בשימוש בשיטות וכלים לעיבוד, ניתוח והדמיה של נתונים. כמו כן, לפתח מיומנויות להפקת תובנות משמעותיות מנתונים מורכבים בכימיה ומדע החומרים בתחומים מגוונים. הכלים יילמדו בשפת התוכנה פייתון. הקורס מהווה בסיס לקורסים מתקדמים נוספים בלימוד מכונה ומדעי נתונים.

תוצרי למידה – יכולות מתקדמות בעיבוד, ניתוח והצגת נתונים עבור מגוון בעיות בכימיה בעזרת שפת תוכנה פייתון, כולל: לנתח ולעבד נתונים נומריים ולהציגם ברמה המתאימה לפרסומים מדעיים, לנהל ביג דאטה, לעשות אנליזה סטטיסטית של נתונים ולהציגם כולל גרסיות, לבצע למידת מכונה בסיסית מונחית ובלתי מונחית, להשתמש בכלים ייחודיים לכמאינפורמטיקה ולהשתמש בכלים ייחודיים למדע החומרים. בסוף הקורס הסטודנטים יוכלו להשתלב בקורסי המשך בנושא למידת מכונה הנלמדים במחלקות אחרות.

### ב. תוכן הקורס:

**רציונל, נושאים:** להעניק לסטודנטים למסלול מדע הנתונים (שנה ג) ותארים מתקדמים במחלקה לכימיה יכולות מתקדמות בניתוח, עיבוד והצגה של נתונים בעזרת שפת תכנות פייתון. כיום כמות המידע שנוצר בניסויים במעבדה ובסימולציות ממוחשבות הולכת וגדלה, וחשוב שלסטודנטים תהיה יכולת לנתח, לעבד ולהציג את הנתונים באופן עצמאי. הנושאים כוללים ניתוח, עיבוד והצגה של נתונים בעזרת ספריות מתקדמות בשפת תכנות פייתון, וכמאינפורמטיקה ואינפורמטיקה במדע החומרים.

מהלך השיעורים: שיעורים פרונטליים כוללים עבודה עצמאית על מחשב, תרגולים הכוללים הדרכה ועבודה עצמאית על מחשב.

תכנית הוראה מפורטת לכל השיעורים: רשימת נושאים / תוכנית הלימודים בקורס – רשימה מפורטת של נושאי ההרצאות בסדר כרונולוגי (רשימה או טבלה כדוגמת המצ"ב)

מס' השיעור	נושא השיעור	קריאה נדרשת	הערות
1	<ul style="list-style-type: none"> <li>- הקדמה לעיבוד נתונים בפיתון</li> <li>- חזרה על תכנות בסיסית בפיתון</li> <li>- מושגים בסיסיים בכמאינפורמטיקה</li> <li>- הכרות עם ספריית RDKit</li> <li>- הכרות עם IPython (IPython Shell and Jupyter Notebook, IPython Magic Commands, Input and Output, Shell Commands, Errors and Debugging, Profiling and Timing code)</li> <li>- הכרות עם PyCHARM</li> </ul>		
2	<ul style="list-style-type: none"> <li>NumPy (The ndarray object, Universal Functions, Aggregations, Working with data files/bases)</li> </ul>		

		NumPy (Vectorization and Broadcasting, Comparisons, masks, and Boolean Logic, Fancy indexing, NumPy in chemistry)	3
		Pandas (Tabular datasets, The Series and DataFrame objects, Operating on data, Handling Missing Data, Aggregation and Grouping, Combining Datasets, Pivot tables, Pandas in chemistry)	4
		MatPlotLib (Line Plots, Scatter Plots, Density and Contour Plots, Histograms and Binning, Figures and Subplots, Colors, Markers, and Line Styles, Text and Annotation, 3D Plots, MatPlotLib in chemistry),	5

		PlotLy, Dash	
		Basic statistics with SciPy/NumPy (Histograms, mean, median, variance, covariance, correlation, physical distributions, probability density function, cumulative distribution functions, regression)	6
		Basic machine learning with Scikit-learn (supervised: k-Nearest Neighbors, decision trees, random forest)	7
		Basic machine learning with Scikit-learn (unsupervised: dendrograms, k-means clustering)	8
		Basic machine learning with Scikit-learn (Principal Components Analysis, t-	9

		distributed Stochastic Neighbor Embedding)	
		Cheminformatics using RDkit 1 (molecular representations, 1D/2D,3D descriptions of molecules)	10
		Cheminformatics using RDkit 2 (molecular descriptors and ML techniques)	11
		Materials Informatics with MatMiner 1	12
		Informatics with MatMiner 2	13

**ג. דרישות קדם:**

קורס "מבוא למחשבים לכימיה" 84190 או קורס מקביל בסיסי בפיתוח.

**ד. חובות / דרישות / מטלות:**

בקורס זה יש חובת נוכחות בהרצאות (85%) ותרגילים (90%).  
יש חובת הגשה של עבודות בית (קודים בפיתוח) ועבודה מסכמת.  
תלמיד שייעדר מעל שתי הרצאות או תרגיל אחד ללא סיבה מוצדקת לא יורשה להגיש  
עבודה מסכמת, ולא יקבל ציון בקורס.

**ה. מרכיבי הציון הסופי:**

50% עבודה מסכמת, 50% תרגילים

ו. ביבליוגרפיה:

חומרים המחולקים להרצאות ותרגילים.

ז. שם הקורס באנגלית:

**Data Processing, Analysis and Visualization in Chemistry**