

მონაცემთა ანალიტიკა - Python

კურსის სახელწოდება	მონაცემთა ანალიტიკა - Python
ინსტრუქტორი	ნოდარ ოქროშიაშვილი
Email	nodar.okroshiashvili@btu.edu.ge
კურსის აღწერა	კურსი მიზნად ისახავს სტუდენტებს შეასწავლოს მონაცემთა ანალიზი პითონის საშუალებით. გააცნოს მონაცემთა ანალიზისთვის საჭირო გავრცელებული მეთოდები და ასწავლოს მათი გამოყენება რეალური ამოცანების გადასაჭრელად.
კურსის შედეგი	კურსის გავლის შემდეგ სტუდენტს შეეძლება: <ul style="list-style-type: none">• კომპანიის/დამსაქმებლის წინაშე მდგარი პრობლემის იდენტიფიცირება.• საჭირო და შესაბამისი მონაცემების მოძიება და ანალიზისათვის მომზადება.• მონაცემების გაწმენდა და დამუშავება. შედეგების ანალიზი.• მონაცემებზე დაყრდნობით კანონზომიერების შესწავლა და მისი გამოყენება გადაწყვეტილების მისაღებად.• მანქანური სწავლების სხვადასხვა ალგორითმების გამოყენება რეგრესიისა და კლასიფიკაციის ამოცანის გადასაჭრელად.
ლექციების განრიგი	TBA
კურსის წინაპირობა	<ul style="list-style-type: none">• Python I და Python II• სტატისტიკა I
სავალდებულო ლიტერატურა	Python for Data Analysis , Wes McKinney, O'Reilly Media, Inc., 2007 Python Data Analytics With Pandas, NumPy, and Matplotlib , Fabio Nelli, Apress Inc, 2018
დამატებითი ლიტერატურა	Python Data Analysis , Ivan Idris, Packt Publishing, 2014

კურსის შინაარსი

ლექცია	მეცადინეობის ფორმა	თემა/აქტივობა	ლიტერატურა
1	ლექცია - 1სთ.	შესავალი <ul style="list-style-type: none"> • მონაცემთა მეცნიერება • მონაცემთა ანალიტიკა • მანქანური სწავლება • ღრმა სწავლება • მონაცემთა ანალიტიკის/მეცნიერების გამოყენება • რატომ Python მონაცემთა ანალიზისათვის? • სილაბუსის გაცნობა • მნიშვნელოვანი ბიბლიოთეკები • სამუშაო გარემოს გაცნობა 	<ul style="list-style-type: none"> • McKinney, გვ 4-8, 15-46, 51-84 • McKinney, გვ 15-46 • Nelli, გვ 103-121
	პრაქტიკული 1სთ-1.5 სთ.	<p>სამუშაო გარემოს მოწყობა და საჭირო ბიბლიოთეკების დაინსტალირება.</p> <p>მონაცემების წაკითხვა ფაილიდან და ფაილში ჩაწერა.</p> <p>მონაცემების ზომის და ტიპების შემოწმება, აღწერითი სტატისტიკა, უნიკალური დაკვირვებები, დაკვირვებების დათვლა.</p>	
2	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Introduction to Pandas <ul style="list-style-type: none"> • Pandas Series • Pandas DataFrame • Index, Selection, Filtering • Re-indexing, Dropping Entries, Axis • Sorting and Ranking • Column rename, reorder, insertion, deletion • Arithmetic operations • Statistics with DataFrame 	<ul style="list-style-type: none"> • McKinney, გვ 123-167 • Nelli, გვ 63-95 • 10 minutes to Pandas • Seeing Theory
3	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Introduction to NumPy <ul style="list-style-type: none"> • Scalar, Vector, Matrix, NdArray • Indexing and Slicing • Boolean Indexing • Universal Functions • Sorting • Statistical methods • Linear Algebra 	<ul style="list-style-type: none"> • McKinney, გვ 85-122 • Nelli, გვ 35-61 • Idris, გვ 25-58, 59-83
4	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Data Cleaning and Preparation <ul style="list-style-type: none"> • Finding and filling missing values • Removing duplicate values • Replacing values • Discretization and Binning • Detecting outliers 	<ul style="list-style-type: none"> • McKinney, გვ 191-219

		<ul style="list-style-type: none"> String manipulations Variable Transformation 	
5	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Join, Combine, Reshape DataFrame <ul style="list-style-type: none"> Combining and Merging Joining and Concatenation Reshaping and Pivoting Groupby Cross Tabulation Long to Wide format Wide to Long format 	<ul style="list-style-type: none"> McKinney, გვ 221-251, 287-316 Nelli, 131-165
6	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Data Visualization <ul style="list-style-type: none"> Line Plots Bar Charts Scatter Plots Histogram Pie Chart Box Plot The Datazaurus Dozen Anscombe's Quartet Unstructured Quartet 	<ul style="list-style-type: none"> McKinney, გვ 253-286 Nelli, გვ 167-235 Matplotlib Tutorials PyPlot Tutorial Example Gallery SeaBorn Plotly Gallery Datazaurus Dozen Anscombe's Quartet
7	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Applied Project 1 <ul style="list-style-type: none"> Fully fledged data analysis Presenting results and insights Discussion about improvements Graphical User Interface for Pandas 	<ul style="list-style-type: none"> Instructor's notebook
8	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Time Series <ul style="list-style-type: none"> Date and Time data types Dealing with datetime objects Indexing, Selection, Sub-setting Periods and Period Arithmetic Converting Timestamp to period and back Resampling Time Shifting Moving Average 	<ul style="list-style-type: none"> McKinney, გვ 317-362
9	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Regression <ul style="list-style-type: none"> One variable regression Decision tree regression Random Forest regression 	<ul style="list-style-type: none"> Instructor's notebook Ordinary Least Squares Decision Tree Regressor Random Forest Regressor

10	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Classification <ul style="list-style-type: none"> • Logistic Regression • Decision Tree classification • Random Forest classification 	<ul style="list-style-type: none"> • Instructor's notebook • <u>Logistic Regression</u> • <u>Decision Tree Classifier</u> • <u>Random Forest Classifier</u>
11	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Applied Project 2 <ul style="list-style-type: none"> • Real world problem to solve • Regression • Classification • Model performance assessment • Room for improvements 	<ul style="list-style-type: none"> • Instructor's notebook
12	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Association Analysis <ul style="list-style-type: none"> • Introduction to association analysis • Analyzing customer basket • Apriori Algorithm 	<ul style="list-style-type: none"> • Instructor's notebook • <u>Market Basket Analysis</u> • <u>Apriori Algorithm</u>
13	ლექცია და პრაქტიკული მეცადინეობა - 2 სთ.	Sentiment Analysis – If time permit <ul style="list-style-type: none"> • Motivation • Summary of Natural Language Processing • Processing textual data • Classifying positive and negative sentiments 	<ul style="list-style-type: none"> • Instructor's notebook