Course Title	Introduction to Data Science	Instructor(s)	Baiko Sai （実務経験のある教員）
		E-mail	
Class Style	Lecture, seminar, etc.	Office Hours	Wednesday PM
Track		Mode of Instruction	Of-line
Credits	2	Allocated Year	Second or third year
Active Learning	Included	Compulsory or Elective	
Course Overview	This class will provide an overview of the basics of data science. Currently, society is flooded with a huge amount of data called big data, but not only data scientists who can analyze and analyze it, but also people who can utilize it while having various discussions with data scientists. The existence of "skilled end users" is important. From this perspective, we will comprehensively discuss the necessary matters.		
Course Objectives	The purpose of this class is to equip the students with the knowledge that will enable them to fully understand the term "data scientist" once they enter the workforce and join a company. From this perspective, the goal is for students to understand ``what data science is generally.''		
Prerequisite			

Course Schedule

No	Contents	Homework
1	Part 1: Data and data analysis #1: What is data analysis? The need for data analysis.	No
2	Part 2: Vector representation and collection of data #2: Vector representation of data, collection of data.	No
3	Part 3: Average, variance, and correlation #3: What are mean, variance, and correlation? Representative values (mean, median, mode) and standard deviation.	No
4	Part 4: Distance and similarity between data #4: Explain the concepts and uses of a huge number of "distances" such as Euclidean distance, Manhattan distance, MAX distance, Mahalanobis distance, discrete distance, Hamming distance, and edit distance.	No
5	Part 5: Clustering and anomaly detection #5: Explain the basic idea of clustering, the k-means method, and other clustering methods. Explain how it is used for anomaly detection.	No
6	Part 6: Fundamentals of data analysis based on linear algebra #6: Explanation of the basic idea of "analysis", vector de-	No

		composition and composition, and "bases" that give compact analysis results.	
	7	Part 7: Principal component analysis #7: Understand the true dimension, the principle of principal component analysis, and the behavior of principal component analysis using a facial image data set as an example.	No
	8	Part 8: Prediction and regression analysis #8: Introducing "prediction" using data, prediction by regression, "model fitting" methods, and multiple regression analysis methods.	No
	9	Part 9: Visualization #9: What is visualization? A technique used in basic visualization.	No
	10	Part 10: Probability and probability distribution #10: What is probability/probability distribution? Explanation of normal distribution and multidimensional normal distribution.	No
	11	Part 11: Confidence intervals and statistical tests #11: Explain confidence intervals and population variance, and explain the concept and basic procedures of statistical tests.	No
	12	Part 12: Unstructured data analysis #12: What is structured data? What is unstructured data? Overview of image recognition technology.	No
	13	Part 13: Pattern recognition and classification #13: Introduction to various methods of pattern recognition.	No
	14	Part 14: Data collection and bias #14: Recognizing sample selection bias, what is personal data, what is open data and copyright?	No
	15	Part 15: Introduction to artificial intelligence, summary #15: What is artificial intelligence (AI)? We will discuss in an active running format about artificial intelligence that is used around us, machine learning, what AI cannot do, etc., and discuss what we have learned so far.	No
Grading	Quiz 20 % Assignments 30 % Credit validation exam 50% Perform a comprehensive evaluation.		
Textbooks	No		

References	「教養としてのデーターサイエンス」内田誠一　他、講談社 「データーサイエンスの基礎」濱田悦生、講談社
NOTES	A short quiz will be given at the end of the class, and a notebook will be created that summarizes the textbook in an easy-to-understand manner. 半導体メーカで、データーサイエンスの学問関連商品開発、企画、セールス経験18年間。