

<b>STUDY MODULE DESCRIPTION FORM</b>		
Name of the module/subject <b>Data integration and data mining</b>		Code <b>1010335431010335195</b>
Field of study <b>Information Engineering</b>	Profile of study (general academic, practical) <b>(brak)</b>	Year /Semester <b>2 / 3</b>
Elective path/specialty <b>Security of Information Technology (IT)</b>	Subject offered in: <b>polish</b>	Course (compulsory, elective) <b>obligatory</b>
Cycle of study: <b>Second-cycle studies</b>	Form of study (full-time, part-time) <b>part-time</b>	
No. of hours Lecture: <b>16</b> Classes: <b>-</b> Laboratory: <b>8</b> Project/seminars: <b>8</b>		No. of credits <b>5</b>
Status of the course in the study program (Basic, major, other) <b>(brak)</b>		(university-wide, from another field) <b>(brak)</b>
Education areas and fields of science and art <b>technical sciences</b>		ECTS distribution (number and %) <b>5 100%</b>
<b>Responsible for subject / lecturer:</b>  dr hab. Tadeusz Pankowski email: tadeusz.pankowski@put.poznan.pl tel. 607-033-007 Wydział Elektryczny ul. Piotrowo 3A 60-965 Poznań		
<b>Prerequisites in terms of knowledge, skills and social competencies:</b>		
1	<b>Knowledge</b>	Ma uporządkowaną i podbudowaną teoretycznie wiedzę w zakresie baz danych.
2	<b>Skills</b>	Potrafi pozyskiwać informacje z literatury, baz danych i innych źródeł; potrafi integrować uzyskane informacje, dokonywać ich interpretacji, a także wyciągać wnioski oraz formułować i uzasadniać opinie.
3	<b>Social competencies</b>	Ma kompetencje społeczne wynikające z programu szkoły średniej.
<b>Assumptions and objectives of the course:</b> Poznanie metod i wybranych narzędzi przeznaczonych do integracji, analizy i eksploracji danych.		
<b>Study outcomes and reference to the educational results for a field of study</b>		
<b>Knowledge:</b> 1. Ma uporządkowaną i podbudowaną teoretycznie wiedzę w zakresie baz danych, hurtowni danych, analizy i eksploracji danych. - [K_W08]		
<b>Skills:</b> 1. Potrafi pozyskiwać informacje z literatury, baz danych i innych źródeł; potrafi integrować uzyskane informacje, dokonywać ich interpretacji, a także wyciągać wnioski oraz formułować i uzasadniać opinie. - [K_U01] 2. Potrafi zaprojektować oraz zrealizować prosty system integracji danych; potrafi analizować i eksplorować dane - [K_U12] 3. Potrafi - przy formułowaniu i rozwiązywaniu informatycznych zadań inżynierskich - dostrzegać ich aspekty pozatechniczne, ekonomiczne i prawne. - [K_U21]		
<b>Social competencies:</b> 1. Ma świadomość odpowiedzialności za pracę własną oraz gotowość podporządkowania się zasadom pracy w zespole i ponoszenia odpowiedzialności za wspólnie realizowane zadania. - [K_K04]		
<b>Assessment methods of study outcomes</b>		
Egzamin pisemny, zaliczenie zajęć laboratoryjnych i projektowych na podstawie zrealizowanych zadań i projektów.		
<b>Course description</b>		

<p><b>Wykłady</b></p> <p>Znaczenie eksploracji danych (data mining). Eksploracja danych na tle rozwoju nauki i technologii przetwarzania danych. Proces odkrywania wiedzy z danych. Metody integracji danych, modelowania danych wielowymiarowych, budowy hurtowni danych. Język MDX i jego znaczenie w przetwarzaniu danych wielowymiarowych. Drzewa decyzyjne jako metoda klasyfikacji obiektów. Wybór zbioru treningowego, problem analizy i predykcji za pomocą drzew decyzyjnych, dziedziny zastosowań. Budowa drzewa decyzyjnego: podstawy teorii informacji (entropia układu); analiza informacyjna ? miara ilości informacji. Klasyfikacja metodą Bayesa. Odkrywanie reguł asocjacyjnych. Algorytm a priori ? odkrywanie zbiorów częstych. Grupowanie obiektów, analiza skupień (klastrowanie obiektów). Klasyfikacja metod grupowania. Grupowanie hierarchiczne. Miary odległości między klastrami (grupami). Algorytm k-średnich. Algorytm k-medoidów. Norma kosinusowa i jej znaczenie w systemach wyszukiwania informacji. Boolowskie systemy wyszukiwania informacji. Eksploracja danych tekstowych. Wykorzystanie metod statystycznych do opisu i analizy danych.</p>		
<p><b>Zajęcia laboratoryjne</b></p> <p>Poznanwanie narzędzi integracji i eksploracji danych i zaznajamianie się z ich praktycznym wykorzystaniem. SSIS (SQL Server Integration Services - z tworzeniem pakietów przepływu włączanie); SSAS (SQL Server Analysis Services - język MDX, kostki, wyodrębnianie miar, wymiarów, agregacje); język DMX; SSRS (SQL Server Reporting Services - integracją z Excelem 2007. Wykorzystanie wspomnianych narzędzi do eksploracji danych: budowa drzew decyzyjnych, klasyfikacja obiektów metodą Baysa, odkrywania reguł asocjacyjnych, eksploracja tekstów (text mining).</p>		
<p><b>Zajęcia projektowe</b></p> <p>Celem zajęć projektowych jest pogłębienie wiedzy w zakresie stosowania metod statystycznych do opisu i analizy danych. Wykonywane są projekty w zakresie: statystyki opisowej, badania zgodności rozkładów, estymacji parametrów, wnioskowania statystycznego, analizy wariancji, analizy korelacji i regresji. Wspomniane zadania i projekty realizowane są z wykorzystaniem systemu Statistica i Excel 2010.</p>		
<p><b>Basic bibliography:</b></p> <ol style="list-style-type: none"> <li>1. Sturm J., Hurtownie danych. Microsoft SQL Server. Przewodnik Techniczny, Microsoft Press/APN PROMISE, Warszawa.</li> <li>2. Han J., Kamber M., Data Mining. Concepts and Techniques, Morgan Kaufmann, 2001.</li> <li>3. C. Seidman, Zgłębianie i analiza danych w Microsoft SQL Server 2000. Przewodnik Techniczny, Microsoft Press/APN PROMISE, Warszawa, 2002.</li> <li>4. M. Rabiej, Statystyka z programem Statistica, Helion, Gliwice, 2012.</li> </ol>		
<p><b>Additional bibliography:</b></p> <ol style="list-style-type: none"> <li>1. D. Mendrala, M. Szeliga, Server SQL 2008. Usługi biznesowe, Helion, Gliwice, 2009.</li> <li>2. R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, ACM Press, 1999.</li> </ol>		
<p><b>Result of average student's workload</b></p>		
<p><b>Activity</b></p>		<p><b>Time (working hours)</b></p>
1. Uczestnictwo w wykładach		16
2. Udział w zajęciach laboratoryjnych		8
3. Udział w zajęciach projektowych		8
4. Przygotowanie projektów i ich realizacja		53
5. Studiowanie notatek z zajęć oraz zalecanej literatury przedmiotu		45
<p><b>Student's workload</b></p>		
<p><b>Source of workload</b></p>	<p><b>hours</b></p>	<p><b>ECTS</b></p>
Total workload	125	5
Contact hours	32	2
Practical activities	78	3