Course Syllabus

Title of the course	**Introduction to Data Mining in Finance and Business**		
Title of the Academic Programme	Master's programme, "Finance"		
Type of the course	Optional		
Prerequisites	Basic statistical methods, programming skills are not mandatory but very helpful		
ECTS workload	3		
Total indicative study hours	Directed Study	Self-directed study	Total
	36	78	114
Course Overview	The course is designed for students interested in data analysis problems. It provides theoretical foundations and practical skills related to data analysis, data mining, text mining, graph mining and ontology-based data analysis. The course is focused on practical aspects of data mining methods and their applications in finance and business.		
Intended Learning Outcomes (ILO)	Students, who positively completed the course, should: • know theoretical foundations of contemporary methods and algorithms used in data analysis area, • know the classification of problems, • know advantages and disadvantages various methods a • be able to choose right methods of analysis to a given problem, • be able to design and implement typical schemes of analysis, • interpret the results delivered by data mining methods, • be able to use data mining in decision making problems, • be able to analyze financial data with the use of data mining methods, • know how develop their theoretical knowledge and practical skills related to data analysis area.		
Teaching and Learning Methods	Lectures, seminars, self-guided studies		

Content and Structure of the Course

№	**Topic / Course Chapter**	**Total**	**Directed Study**		**Self-directed Study**
			Lectures	**Tutorials**	
1.	Introduction do data analysis. Data preprocessing. Introduction to R language.	8	2	2	10
2.	Complex data structures in R. Scripts and control statements. Data preprocessing and visualization. Linear regression. Logistic regression	8	2	2	9
3.	Neural network models	8	2	2	8
4.	Decision tree models. Naive Bayes classifier. Support vector machine model. Association rules.	8	2	2	8

5.	Problem of dimensionality reduction. Multidimensional scaling. Correspondence analysis. Genetic algorithms.	8	2	2	8
6.	Cluster analysis.	8	2	2	8
7.	Introduction to text mining.	8	2	2	9
8.	Sentiment analysis and ontology-based analysis of text documents.	8	2	2	9
9.	Network analysis.	8	2	2	9
Total study hours		**114**	**18**	**18**	**78**

Indicative Assessment Methods and Strategy	**The grade (G)** is calculated as follows: $G = 0{,}2 \cdot G_{t1} + 0{,}2 \cdot G_{t2} + 0{,}6 \cdot G_{EX}$, where G_{t1} – grade for test 1 G_{t2} – grade for test 2 **G_{EX} – grade for the final examination**
Readings / Indicative Learning Resources	Mandatory • Bramer M., *Principles of Data Mining*, Second Edition, Springer, 2013 • W. N. Venables, D. M. Smith and the R Core Team, *An Introduction to R*, available at: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf Optional • *listed in the appendix 1: Course Content*

Indicative Self- Study Strategies	**Type**	**+/–**	**Hours**
	Reading for seminars / tutorials (lecture materials, mandatory and optional resources)		**30**
	Assignments for seminars / tutorials / labs		**15**
	E-learning / distance learning (MOOC / LMS)		
	Fieldwork		
	Project work		**20**
	Other (please specify)		
	Preparation for the exam		**13**

Academic Support for the Course	Academic support for the course is provided via LMS, where students can find: guidelines and recommendations for doing the course; guidelines and recommendations for self-study; samples of assessment materials
Facilities, Equipment and Software	• R system (https://cran.r-project.org/) • RStudio system, Desktop version (free version) (https://www.rstudio.com/products/rstudio/download/)
Course Instructor	Pawel Lula, PhD, pawel.lula@post.pl Edward Mailov, edwardmailov@yandex.ru

Intended Learning Outcomes (ILO) Delivering

Programme ILO(s)	Course ILO(s)	Teaching and Learning Methods for delivering ILO(s)	Indicative Assessment Methods of Delivered ILO(s)
ILO_1 Understand the challenges of uncertain economic environment, assess them and take appropriate financial and investment decisions	Students should know how to: evaluate economic and financial environment, identify main determinants of financial and investment decisions, evaluate the impact of economic and financial decisions on real-life phenomena.	Discussions of real-life issues. Individual projects. Additional reading.	Exam Reports In-class discussions
ILO_3 Use strong analytical skills and apply them to solve practical problems	Students should know how to: use formal methods to describe, analyze and predict social, economic and financial phenomena, implement formal models in a form of computer programs, use computer models to analyze real-life situations.	Discussion of real-life cases. Additional reading. Individual projects in R realized at home.	Exam Reports In-class discussion
ILO_4 Examine and critically appraise research methods and tools relevant for research in finance	Students should know how to: choose proper methods and research tools for a given practical problem, choose	Presentations and discussions of real-life issues. Analysis of publications in leading academic journals.	Exam Reports In-class discussions
ILO_9 Demonstrate a range of generic skills including information and time management, team and project work, computing and autonomous learning, digital skills	Students should know how to: use ICT solutions in solving real-life problems, work together with other team members,	Individual projects realized in R. Additional reading. E-learning.	Reports In-class discussion

	develop personal knowledge and skills.		
ILO_{10} Demonstrate an innovative, open and ethical mindset	Students should be ready to: work in multicultural environment, be innovative, accept various world views and systems of values.	Discussions	In-class discussions

Annex 1

Course Content

Lecture 1: Introduction do data analysis. Data preprocessing

1. Introduction to data analysis
2. Confirmatory and exploratory approach in data analysis
3. Data types and sources
4. Data analysis process
5. Types of problems
6. Introduction to R language
7. Simple data types
8. Assignment statement
9. Basic input/output commands
10. Control statements

Reading:

- Bramer M., *Principles of Data Mining*, Second Edition, Springer, 2013 (Chapters: 1, 2)

Seminar 1: Introduction to R language.

1. Command-line mode
2. R-Studio
3. Immediate calculations in R
4. Operators
5. Variables
6. Input/output operations
7. Scripts and control statements

Reading:

- https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

Lecture 2: Complex data structures. Linear regression. Logistic regression

1. Complex data structures (vectors, matrices, data frames, lists)
2. Data preprocessing
3. Data visualization
4. Linear regression
5. Logistic regression

Reading:

Complex data structures:

- https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

Data preprocessing:

- Bramer M., *Principles of Data Mining*, Second Edition, Springer, 2013 (Chapter: 2)

Data visualization:

- https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
- https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/
- https://cran.r-project.org/web/packages/ggplot2/vignettes/extending-ggplot2.html
- http://r4ds.had.co.nz/data-visualisation.html
- https://www.statmethods.net/advgraphs/ggplot2.html

- https://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html#ggplot2_vs_base_graphics
- https://www.rstudio.com/wp-content/uploads/2015/04/ggplot2-cheatsheet.pdf

Linear regression:

- http://r-statistics.co/Linear-Regression.html
- https://www.statmethods.net/stats/regression.html
- https://www.r-bloggers.com/simple-linear-regression-2/

Logistic regression:

- https://www.r-bloggers.com/how-to-perform-a-logistic-regression-in-r/
- https://www.analyticsvidhya.com/blog/2015/11/beginners-guide-on-logistic-regression-in-r/
- https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/logistic-regression-analysis-r/tutorial/

Seminar 2: Scripts and control statements. Data preprocessing and visualization. Linear regression. Logistic regression – practical aspects and applications

1. Complex data structures (vectors, matrices, data frames, sets, lists)
2. Data preprocessing
3. Data visualization
4. Linear regression
5. Logistic regression

Reading:

Complex data structures:

- https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

Data preprocessing:

- Bramer M., *Principles of Data Mining*, Second Edition, Springer, 2013 (Chapter: 2)

Data visualization:

- https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
- https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/
- https://cran.r-project.org/web/packages/ggplot2/vignettes/extending-ggplot2.html
- http://r4ds.had.co.nz/data-visualisation.html
- https://www.statmethods.net/advgraphs/ggplot2.html
- https://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html#ggplot2_vs_base_graphics
- https://www.rstudio.com/wp-content/uploads/2015/04/ggplot2-cheatsheet.pdf

Linear regression:

- http://r-statistics.co/Linear-Regression.html
- https://www.statmethods.net/stats/regression.html
- https://www.r-bloggers.com/simple-linear-regression-2/

Logistic regression:

- https://www.r-bloggers.com/how-to-perform-a-logistic-regression-in-r/
- https://www.analyticsvidhya.com/blog/2015/11/beginners-guide-on-logistic-regression-in-r/
- https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/logistic-regression-analysis-r/tutorial/

Lecture 3: Neural network models

1. Model of artificial neuron

2. Taxonomy of neural network models
3. Multilayer perceptrons
4. RBF networks
5. Kohonen networks
6. Deep learning

Reading:

- https://natureofcode.com/book/chapter-10-neural-networks/
- https://medium.com/datathings/neural-networks-and-backpropagation-explained-in-a-simple-way-f540a3611f5e
- https://towardsdatascience.com/radial-basis-functions-neural-networks-all-we-need-to-know-9a88cc053448
- https://page.mi.fu-berlin.de/rojas/neural/chapter/K15.pdf
- http://www.scholarpedia.org/article/Kohonen_network
- https://www.deeplearningbook.org/

Seminar 3: Practical aspects of neural network model building

1. Choice of proper neural network model and learning algorithm
2. Overtraining problem
3. Classification problem solving with neural network models
4. Optimization of neural network model structure

Reading:

- https://cran.r-project.org/web/packages/RSNNS/
- https://clarkdatalabs.github.io/soms/SOM_NBA
- https://www.jstatsoft.org/article/view/v021i05/v21i05.pdf
- https://ifdo.ca/~seymour/R/neural.pdf
- https://www.datacamp.com/community/tutorials/keras-r-deep-learning

Lecture 4: Decision tree models. Naive Bayes classifier. Support vector machine model. Association rules – methods and algorithms

Topics:

1. Introduction to decision tree models
2. Measurement of group homogeneity (entropy, Gini coefficient)
3. CART (Classification and Regression Tree) model
4. Ensemble models – bagging technique, boosting technique, random forest models
5. Random forests
6. Naive Bayes classifier
7. Support vector machine
8. Association rules and their evaluation
9. Apriori and ECLAT algorithms
10. Market basket analysis and churn analysis

Reading:

Decision trees:

- Bramer M., *Principles of Data Mining*, Second Edition, Springer, 2013 (Chapters: 3 - 14)
- http://www.ccs.miami.edu/~hishwaran/papers/decisionTree_intro_IR2009_EMDM.pdf

Support vector machine:

- https://en.wikipedia.org/wiki/Support_vector_machine
- https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/
- http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf

Naive Bayes classifier:

- https://en.wikipedia.org/wiki/Naive_Bayes_classifier

Association rules:

- Bramer M., *Principles of Data Mining*, Second Edition, Springer, 2013 (Chapters: 16 - 18)

Seminar 4: Decision tree models. Naive Bayes classifier. Support vector machine model. Association rules – practical aspects and applications

Topics:

1. Decision tree building process
2. Decision tree visualization
3. CART algorithm
4. Bagging and boosting techniques
5. Random forest models
6. Naive Bayes classifier
7. Support vector machine
8. Identification of associate rules
9. Apriori algorithm
10. ECLAT algorithm
11. Market basket analysis

Reading:

12. https://www.datacamp.com/community/tutorials/decision-trees-R
13. https://www.guru99.com/r-decision-trees.html
14. http://ijcsit.com/docs/aceit-conference-2016/aceit201639.pdf
15. https://cran.r-project.org/web/packages/rpart.plot/index.html
16. https://cran.r-project.org/web/packages/rpart/index.html
17. https://cran.r-project.org/web/packages/adabag/index.html
18. https://www.jstatsoft.org/article/view/v054i02
19. https://cran.r-project.org/web/packages/randomForest/
20. https://datascienceplus.com/random-forests-in-r/
21. https://cran.r-project.org/web/packages/arules/index.html
22. https://www.jstatsoft.org/article/view/v015i09/v15i09.pdf

Lecture 5: Problem of dimensionality reduction. Multidimensional scaling. Correspondence analysis. Genetic algorithms – methods and algorithms

Topics:

1. Dimensionality reduction problem in data analysis
2. Principal component analysis
3. Singular Value Decomposition and its application
4. Multidimensional scaling
5. Correspondence analysis
6. Genetic algorithms and its application in dimensionality reduction

Reading:

- http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
- https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/
- http://infolab.stanford.edu/~ullman/mmds/ch11.pdf
- http://www.di.fc.ul.pt/~jpn/r/svd/svd.html
- https://www.displayr.com/singular-value-decomposition-in-r/
- https://arxiv.org/ftp/arxiv/papers/1403/1403.2877.pdf
- https://www.math.uwaterloo.ca/~aghodsib/courses/f06stat890/readings/tutorial_stat890.pdf

- http://www.statisticshowto.com/multidimensional-scaling/
- https://www.utdallas.edu/~herve/Abdi-MDS2007-pretty.pdf
- https://www.displayr.com/how-correspondence-analysis-works/
- https://www.analyticsvidhya.com/blog/2017/07/introduction-to-genetic-algorithm/

Seminar 5: Problem of dimensional reduction. Multidimensional scaling. Correspondence analysis. Genetic algorithms – practical aspects and applications

Topics:

1. PCA algorithm
2. SVD as a tool for dimensional reduction
3. Multidimensional scaling
4. Correspondence analysis
5. GA as a tool for variable prediction

Reading:

- http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
- http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/113-ca-correspondence-analysis-in-r-essentials/
- https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/
- http://infolab.stanford.edu/~ullman/mmds/ch11.pdf
- http://www.di.fc.ul.pt/~jpn/r/svd/svd.html
- https://www.displayr.com/singular-value-decomposition-in-r/
- https://arxiv.org/ftp/arxiv/papers/1403/1403.2877.pdf
- https://www.math.uwaterloo.ca/~aghodsib/courses/f06stat890/readings/tutorial_stat890.pdf
- http://www.statisticshowto.com/multidimensional-scaling/
- https://www.utdallas.edu/~herve/Abdi-MDS2007-pretty.pdf
- https://www.displayr.com/how-correspondence-analysis-works/
- https://www.analyticsvidhya.com/blog/2017/07/introduction-to-genetic-algorithm/
- https://docs.google.com/viewer?url=http%3A%2F%2Fwww.mmds.org%2Fmmds%2Fv2.1%2Fch11-dimred.pptx
- https://cran.r-project.org/web/packages/GA/vignettes/GA.html
- https://www.jstatsoft.org/article/view/v053i04
- https://www.r-bloggers.com/genetic-algorithms-a-simple-r-example/

Lecture 6: Cluster analysis – methods and algorithms

Topics:

1. Hierarchical methods
2. *k-means* method
3. Comparison of clustering process results
4. Evaluation of clustering quality
5. Model-based clustering

Reading:

- Bramer M., *Principles of Data Mining*, Second Edition, Springer, 2013 (Chapter: 19)
- https://www-users.cs.umn.edu/~kumar001/dmbook/ch8.pdf
- http://www.yorku.ca/ptryfos/f1500.pdf
- https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/
- https://www.sciencedirect.com/science/article/pii/0377042787901257
- https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68
- https://hal.inria.fr/hal-01252671/document

- https://hal.archives-ouvertes.fr/hal-01252673v2/document
- https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm
- https://arxiv.org/pdf/1807.01987.pdf

Seminar 6: Cluster analysis – practical aspects and applications

Topics:

1. Hierarchical methods
2. *k-means* method
3. Comparison of clustering process results
4. Evaluation of clustering quality
5. Model-based clustering

Reading:

- https://uc-r.github.io/kmeans_clustering
- https://www.statmethods.net/advstats/cluster.html
- https://cran.r-project.org/web/packages/cluster/
- http://www.iasri.res.in/ebook/win_school_aa/notes/Cluster_Analysis_usingR.pdf
- https://cran.r-project.org/web/packages/mclust/vignettes/mclust.html
- https://data-flair.training/blogs/r-clustering-tutorial/
- http://cc.oulu.fi/~jarioksa/opetus/metodi/sessio3.pdf
- http://www.instantr.com/2013/02/12/performing-a-cluster-analysis-in-r/

Lecture 7: Introduction to text mining – methods and algorithms

Topics:

1. Document preprocessing
2. Frequency matrix and its analysis
3. Similarity of documents and words
4. Latent Semantic analysis
5. Latent Dirichlet Allocation

Reading:

- Bramer M., *Principles of Data Mining*, Second Edition, Springer, 2013 (Chapter: 20)
- http://people.ischool.berkeley.edu/~hearst/text-mining.html
- http://people.ischool.berkeley.edu/~hearst/papers/acl99/acl99-tdm.html
- https://www.cms.waikato.ac.nz/~ihw/papers/04-IHW-Textmining.pdf
- http://lsa.colorado.edu/papers/dp1.LSAintro.pdf
- http://webhome.cs.uvic.ca/~thomo/svd.pdf
- http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Seminar 7: Introduction to text mining – practical aspects and applications

Topics:

1. Document preprocessing
2. Frequency matrix and its analysis
3. Similarity of documents and words
4. Latent Semantic analysis
5. Latent Dirichlet Allocation

Reading:

- https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf
- http://kenbenoit.net/pdfs/text_analysis_in_R.pdf
- https://rpubs.com/pjmurphy/265713
- https://eight2late.wordpress.com/2015/05/27/a-gentle-introduction-to-text-mining-using-r/

- https://tutorials.quanteda.io/
- https://www.tidytextmining.com/topicmodeling.html

Lecture 8: Sentiment analysis and ontology-based analysis of text documents – methods and algorithms

Topics:

1. Sentiment analysis
2. Ontologies in data analysis
3. Ontology-bases similarity
4. Ontology-based analysis of text documents

Reading:

- https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17
- https://www.sciencedirect.com/science/article/pii/S2090447914000550
- https://file.scirp.org/pdf/JCC_2017020917284790.pdf
- http://ijirae.com/images/downloads/vol1issue2/ACS10115.April14.28.pdf
- https://pdfs.semanticscholar.org/b3b3/9fdc57b869cf30ed3f6f0e22277c1008fc2b.pdf

Seminar 8: Sentiment analysis and ontology-based analysis of text documents – practical aspects and applications

Topics:

1. Sentiment analysis
2. Ontologies in data analysis
3. Ontology-bases similarity
4. Ontology-based analysis of text documents

Reading:

- https://cran.r-project.org/web/packages/SentimentAnalysis/vignettes/SentimentAnalysis.html
- https://datascienceplus.com/sentiment-analysis-with-machine-learning-in-r/
- https://towardsdatascience.com/sentiment-analysis-in-r-good-vs-not-good-handling-negations-2404ec9ff2ae
- https://cran.r-project.org/web/packages/ontologySimilarity/index.html
- https://cran.r-project.org/web/packages/ontologyIndex/index.html

Lecture 9: Network analysis – methods and algorithms

Topics:

1. Network structure analysis
2. Centrality
3. Network visualization
4. Link analysis
5. Bipartite graphs
6. Similarity of graphs
7. Frequent subgraph mining

Reading:

- http://courses.washington.edu/ir2010/readings/butts.pdf
- http://www.mjdenny.com/workshops/SN_Theory_I.pdf
- https://en.wikipedia.org/wiki/Network_theory
- https://en.wikipedia.org/wiki/Social_network_analysis
- https://www.analyticsvidhya.com/blog/2018/04/introduction-to-graph-theory-network-analysis-python-codes/
- http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/07/ECPR01.pdf

- https://www.uva.nl/binaries/content/documents/personalpages/n/o/w.denooy/en/tab-one/tab-one/cpitem%5B26%5D/asset?1355372751494
- https://core.ac.uk/download/pdf/82599177.pdf
- https://airccj.org/CSCP/vol2/csit2117.pdf

Seminar 9: Network analysis – practical aspects and applications

Topics:

1. Network structure analysis
2. Centrality
3. Network visualization
4. Link analysis
5. Bipartite graphs
6. Similarity of graphs
7. Frequent subgraph mining

Reading:

8. http://pablobarbera.com/big-data-upf/html/02a-networks-intro-visualization.html
9. http://kateto.net/networks-r-igraph
10. https://www.csc2.ncsu.edu/faculty/nfsamato/practical-graph-mining-with-R/slides/pdf/Frequent_Subgraph_Mining.pdf

Annex 2

Assessment Methods and Criteria

Assessment Methods

Types of Assessment	Forms of Assessment	Modules			
		1	2	3	4
Formative Assessment	Test	*			
	Essay				
	Report/Presentation	*			
	Project	*			
	In-class Participation	*			
	Other (write appropriate control forms for the course)				
Interim Assessment (if required)	Assignment (e.g. written assignment)				
Summative Assessment	Exam	*			

Assessment Criteria

In-class Participation

Grades	Assessment Criteria
«Excellent» (8-10)	A critical analysis which demonstrates original thinking and shows strong evidence of preparatory research and broad background knowledge.
«Good» (6-7)	Shows strong evidence of preparatory research and broad background knowledge. Excellent oral expression.
«Satisfactory» (4-5)	Satisfactory overall, showing a fair knowledge of the topic, a reasonable standard of expression. Some hesitation in answering follow-up questions and/or gives incomplete or partly irrelevant answers.
«Fail» (0-2)	Limited evidence of relevant knowledge and an attempt to address the topic. Unable to offer relevant information or opinion in answer to follow-up questions.

Project Work

Grades	Assessment Criteria
«Excellent» (8-10)	A well-structured, analytical presentation of project work. Shows strong evidence and broad background knowledge. In a group presentation all members contribute equally and each contribution builds on the previous one clearly; Answers to follow-up questions reveal a good range and depth of knowledge beyond that covered in the presentation and show confidence in discussion.
«Good» (6-7)	Clearly organized analysis, showing evidence of a good overall knowledge of the topic. The presenter of the project work highlights key points and responds to follow up questions appropriately. In group presentations there is evidence that the group has met to discuss the topic and is presenting the results of that discussion, in an order previously agreed.
«Satisfactory» (4-5)	Takes a very basic approach to the topic, using broadly appropriate material but lacking focus. The presentation of project work is largely unstructured, and some points are irrelevant to the topic. Knowledge of the topic is limited and there may be evidence of basic misunderstanding. In a group presentation, most of the work is done by one or two students and the individual contributions do not add up.
«Fail» (0-2)	Fails to demonstrate any appropriate knowledge.

Written Assignments (Essay, Test/Quiz, Written Exam, etc.)

Grades	Assessment Criteria
«Excellent» (8-10)	Has a clear argument, which addresses the topic and responds effectively to all aspects of the task. Fully satisfies all the requirements of the task; rare minor errors occur;
«Good» (6-7)	Responds to most aspects of the topic with a clear, explicit argument. Covers the requirements of the task; may produce occasional errors.
«Satisfactory» (4-5)	Generally addresses the task; the format may be inappropriate in places; display little evidence of (depending on the assignment): independent thought and critical judgement include a partial superficial coverage of the key issues, lack critical analysis, may make frequent errors.
«Fail» (0-2)	Fails to demonstrate any appropriate knowledge.

Annex 3

Description of assignments

Assignment	Description
Test 1	Test is hand written and delivered on print outs. Total number of points is 20 if test is finished during 25 minutes. Block A includes 3 multiple choice questions and 2 open ended (2 points each). Block B includes 2 practical problems (5 points each) Example: **Block A** [10% total, 2% each] <u>Select the best answer(s):</u> 1) Which control statements can we use while creating data analysis programs in R in order to **repeat** the particular **operations** for some time? a) If e) Break b) While f) Until c) As g) Repeat d) For h) Next 2) Main **purposes of regression** modelling are: a) testing of hypotheses about relationship between predictor and response b) testing of relationship among dependent variables c) testing of relationship among independent variables d) building of predictive models e) building of evaluative models f) building of linear equation connecting Y with X 3) <u>Answer briefly but at the same time accurately on the questions below:</u> 4) Our course is about mining of the data. What is in your opinion the difference among Data, Information and Knowledge? Which one is more important for the support of decision making? 5) **Block B** [10% total, 5% each] <u>Solve the following practical problems</u> 1) Could you explain what will be displayed as the result and intuitively interpret the numbers age <- seq (10:20) print(mean(age)) print(median(age)) print(sd(age)) 2) Which of the following results of linear regression model building will help you to evaluate whether the model is reliable, in other words its fitness. Give as many explanations as you can.

	<pre>Coefficients:
	Estimate Std. Error t value Pr(>|t|)
	(Intercept) -1.446e+03 1.042e+02 -13.871 <2e-16 ***
	cylinders -3.299e-01 3.321e-01 -0.993 0.321
	displacement 7.678e-03 7.358e-03 1.044 0.297
	horsepower -3.914e-04 1.384e-02 -0.028 0.977
	weight -6.795e-03 6.700e-04 -10.141 <2e-16 ***
	acceleration 8.527e-02 1.020e-01 0.836 0.404
	modelyear 7.534e-01 5.262e-02 14.318 <2e-16 ***
	--- 	
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
	
	Residual standard error: 3.435 on 385 degrees of freedom
	Multiple R-squared: 0.8093, Adjusted R-squared: 0.8063
	F-statistic: 272.2 on 6 and 385 DF, p-value: < 2.2e-16</pre>

Test 2	Test is performed in the form of creating program in R language using R studio software. Example: Create the program which allows to input the number from the keyboard and then display its factorial (hint: use "for" statement)
Exam	Exam is provided in a form of final practical project and contributes up to 60% of the final grade. The main task of the project is to apply one of the data analysis methods learned during the course on the selected data set. Report is submitted in the form of 5-7 pages word file describing initial data set and explaining the results of the performed analysis. More details are provided on the first seminar.