Ministry of Science and Higher Education of the Russian Federation
NATIONAL RESEARCH
TOMSK STATE UNIVERSITY (NR TSU)

Institute of Applied Mathematics and Computer Science

APPROVE:
Director

_____A. V. Zamyatin

Work program of the discipline

Data Visualization

in the major of training

01.04.02 Applied mathematics and informatics

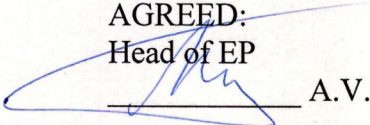Orientation (profile) of training:

Big Data and Data Science

Form of study
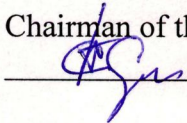**full-time**

Qualification
Master

Year of admission
2023

Code of discipline in the curriculum: B1.O.09

AGREED:
Head of EP

_____ A.V. Zamyatin

Chairman of the EMC

_____ S.P. Sushchenko

Tomsk – 2023

1. Purpose and planned results of mastering the discipline

The purpose of mastering the discipline is the formation of the following competencies:
- PC-1 - the ability to develop and apply mathematical methods, algorithms, software for solving problems of research and design activities.
The results of mastering the discipline are the following indicators of the achievement of competencies:
IPK-1.3 Develops new methods, models, algorithms and software for solving problems in the field of professional activity.

2. Tasks of mastering the discipline

– to form skills for solving applied professional problems using methods and means of data and knowledge visualization;
- to teach to analyze the tasks of professional activity by means of information technology;
– to teach how to interpret visualization results and describe them correctly;
– to form the skills of using libraries of modern scripting languages (Python, R) for data and knowledge visualization.

3. The place of discipline in the structure of the educational program

Discipline belongs to the mandatory part of the educational program.

4. Semester of mastering and form of intermediate certification in the discipline

Second semester, credit.

5. Entrance requirements for mastering the discipline

For the successful mastering of the discipline, training outcomes are required in the following disciplines: "Introduction to Data Science & Data Mining".

6. Implementation language

English.

7. Scope of discipline

The total labor intensity of the discipline is 3 credits, 108 hours, of which:
- lectures: 10 hours
- laboratory: 20 hours
  including practical training: 0 h.
The volume of independent work of the student is determined by the curriculum.

8. The content of the discipline, structured by topics

Topic 1. Introduction to Visualization with Python – Basic and Customized Plotting. Reading Data from Files, Observing and Describing Data, Plotting with pandas and seaborn, Creating Simple Plots to Visualize a Distribution of Variables, Plotting and Analyzing a Histogram, Creating a Bar Plot and Calculating the Mean Price Distribution, Creating Bar Plots Grouped by a Specific Feature, Tweaking the Plot Parameters of a Grouped Bar Plot, Annotating a Bar Plot.
Lab 1. Basic and Customized Plotting
Activity 1: Analyzing Different Scenarios and Generating the Appropriate Visualization

Topic 2. Static Visualization – Global Patterns and Summary Statistics.

Creating Plots that Present Global Patterns in Data, Scatter Plots, Creating a Static Scatter Plot, Creating a Static Hexagonal Binning Plot, Creating a Static Contour Plot, Creating a Static Line Plot, Presenting Data across Time with multiple Line Plots. Creating and Exploring a Static Heatmap, Creating Plots That Present Summary Statistics of Your Data, Creating and Exploring a Static Box Plot, Creating a Static Violin Plot).

Lab 2. Static Visualization
Activity 2: Design Static Visualization to Present Global Patterns and Summary Statistics

Topic 3. From Static to Interactive Visualization (Static versus Interactive Visualization , Applications of Interactive Data Visualizations, Getting Started with Interactive Data Visualizations, Interactive Data Visualization with Plotly Express, Creating an Interactive Scatter Plot.

Lab 3. Interactive Visualization
Activity 3: Creating Different Interactive Visualizations Using Plotly Express

Topic 4. Interactive Visualization of Geographical Data (Choropleth Maps, Worldwide Choropleth Maps, Creating a Worldwide Choropleth Map, Adding Animation to a Choropleth, Plots on Geographical Maps, Creating a Scatter Plot on a Geographical Map, Creating a Bubble Plot on a Geographical Map, Line Plots on Geographical Maps. Creating Line Plots on a Geographical Map)

Lab 4. Interactive Visualization of Geographical Data
Activity 4: Creating a Choropleth Map to Represent Total Renewable Energy Production and Consumption across the World.

9. Ongoing evaluation

The ongoing evaluation is carried out by monitoring attendance, conducting tests, tests on lecture material, performing laboratory work, and is recorded in the form of a checkpoint at least once a semester.

10. The procedure for conducting and criteria for evaluating the intermediate certification

The test with an assessment is carried using computer. The task contains 3 practical tasks. Example:
Final task. Using the data set given by your teacher, create two free-form graphs showing your ability to use different settings for the data set. Give a description of the data set and what the graph shown shows.

The results of the assessment with an assessment are determined by the marks "excellent", "good", "satisfactory", "unsatisfactory".

The current control takes into account the student's performance of laboratory work and answers to questions at colloquia. The points scored are taken into account during the intermediate certification. Grades "excellent", "good", "satisfactory", "unsatisfactory" are given with the number of points scored: 90-100, 70-89, 55-69 and 0-54, respectively

11. Educational and methodological support

a) Electronic training course on the discipline at the electronic university "Moodle" - https://moodle.tsu.ru/course/view.php?id=5464
b) Evaluation materials of the current control and intermediate certification in the discipline (Appendix 1).
c) Plan of seminars / practical classes in the discipline.
Not provided

d) Guidelines for laboratory work.

The implementation of laboratory work is provided for each studied topic of the discipline. Methodical manuals are placed in the IPMCN local network in the X:\Workspace\Master\Knowledge Representation and Data Visualization directory.

e) Guidelines for organizing students' independent work.

To strengthen the studied material, it is planned to carry out an individual project within hours of independent work. The project can be completed both individually and in a mini-group (2-3 people), provided that the amount of work will also be increased. At the end of the semester, a mini-presentation on the results of the work is presented for each project.

12. List of educational literature and Internet resources

a) basic literature:

- Eric Matiz. Learning Python: game programming, data visualization, web applications. - St. Petersburg: Peter, 2021. - 587 p.

– McKinney, W. Python and data analysis. – M.: DMK Press, 2020. – 540 p.

– Mastitsky, S.E., Shitikov V.K. Statistical analysis and data visualization using R. - DMK Press, 2015. - 496 p.

– Mastitsky, S.E. Data visualization with ggplot2. – M.: DMK Press, 2017. – 222 p.

B) Internet resources:

Information and reference systems:

–　　Knowledge　　Representation　　and　　Data　　Visualization https://moodle.ido.tsu.ru/course/view.php?id=1403

– 7 most popular Python libraries for data visualization in Data Science https://medium.com/@bigdataschool/7-самых-популярных-python-библиотек-для-визуализации-данных-в-data-science-c0a4875df93b

– An Overview of Data Visualization Packages in Python https://medium.com/@alibekashirali/обзор-пакетов-для-визуализации-данных-на-python-a1525b2a5cc7

– Data Visualization Catalog https://datavizcatalogue.com/

– Why and how to use data visualization? https://infogra.ru/infographics/zachem-i-kak-ispolzovat-vizualizatsiyu-dannyh

– Open Machine Learning Course https://habr.com/ru/company/ods/blog/323210/

– An overview of methods for visualizing multidimensional data https://cyberleninka.ru/article/n/obzor-metodov-vizualizatsii-mnogomernyh-danny

– How to create a dashboard using Google Sheets https://blog.sheetgo.com/spreadsheets-tips/how-to-create-a-dashboard-in-google-sheets/

Electronic library systems

– TSU Scientific Library https://www.lib.tsu.ru/

– Electronic library system "Lan" https://e.lanbook.com/

– CyberLeninka https://cyberleninka.ru/

13. List of information technologies

a) licensed and freely distributed software:

– Microsoft Office Standard 2013 Russian: software package. Includes applications: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);

- publicly available cloud technologies (Google Docs, Yandex disk, etc.).

– Python3 and R programming tools:

☐Python3

☐RStudio.

b) information reference systems:

– Electronic catalog of the TSU Scientific Library – http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system

– TSU electronic library (repository) – http://vital.lib.tsu.ru/vital/access/manager/Index

– EBS Lan – http://e.lanbook.com/

– EBS Student Advisor – http://www.studentlibrary.ru/

– Urayt educational platform – https://urait.ru/

– EBS ZNANIUM.com – https://znanium.com/

- EBS IPRbooks - http://www.iprbookshop.ru/

c) professional databases:

– Artificial intelligence and its applications. News of the development of quantum computers. Research of artificial neural networks https://ai-news.ru

– Portal of world competitions of DM teams https://www.kaggle.com/

– Machine Learning Repository https://archive.ics.uci.edu/ml/index.php

14. Logistics

Audiences for conducting lecture-type classes.

Audiences for laboratory classes, individual and group consultations, current control and intermediate certification.

Rooms for independent work, equipped with computer technology and access to the Internet, to the electronic information and educational environment and to information reference systems.

For the material and technical support of the discipline, the availability of computer equipment with installed appropriate software and other equipment that supports presentations and access to the Internet is required.

15. Authors information

Marukhina Olga Vladimirovna, Ph.D. tech. Sciences, Associate Professor, Associate Professor of the Department of Theoretical Foundations of Informatics, TSU.