

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN
FEDERATION

Federal State Autonomous Educational Institution for Higher Education
«NORTH-CAUCASUS FEDERAL UNIVERSITY»

METHODICAL INSTRUCTIONS

to perform the laboratory work on the discipline " Multidimensional data analysis"
for the areas of training 09.04.03 "Applied Information Science" (Directivity
(profile) "Knowledge Management")

Stavropol, 2026

CONTENT

INTRODUCTION

Lab number 1. OLAP application architecture

Lab number 2. Using the AnalyzeKeyInfluencers and DetectCategories tools

Lab number 3. Using the "FillFromExample" and "Forecast" tools

Lab number 4. Using the "HighlightExceptions" and "ScenarioAnalysis" tools

Lab number 5. Using the "Prediction Calculator" and "ShoppingbasketAnalysis" tools

Lab number 6. Using Data Mining Client Tools for Excel 2007 to prepare data

Lab number 7. Use Data Mining Client Tools for Excel 2007 to create a mining model

Lab number 8. Analysis of forecast accuracy and use of a mining model

Lab number 9. Build a clustering model

Lab number 10. Creating a data source view

Lab number 11. Creating the structure and model of mining. Clustering task

RECOMMENDED LITERATURE

Introduction

Information systems of a serious enterprise, as a rule, contain applications designed for complex analysis of data, their dynamics, trends, etc. Accordingly, top management becomes the main consumers of the analysis results. Such an analysis, ultimately, is intended to facilitate decision-making. And in order to make any management decision, it is necessary to have the necessary information, usually quantitative. To do this, it is necessary to collect these data from all information systems of the enterprise, lead to a common format and then analyze it. To do this, create a data warehouse (Data Warehouses).

Lab number 1. OLAP application architecture

The goal of the work: to acquire practical skills in analyzing the architecture of OLAP applications.

Formed competences or their parts - GPC-2; PC-2; PC-3

The theoretical part.

Multidimensionality in OLAP applications can be divided into three levels:

The multidimensional data representation is the end-user facility providing multidimensional visualization and data manipulation; the multidimensional representation layer is abstracted from the physical data structure and perceives the data as multidimensional.

Multidimensional processing is a tool (language) for formulating multidimensional queries (the traditional relational SQL language turns out to be unsuitable here) and a processor that can process and execute such a query.

Multidimensional storage is a means of physically organizing data that ensures the efficient execution of multidimensional queries.

The first two levels are necessarily present in all OLAP media. The third level, although it is widespread, is not necessary, since the data for the multidimensional representation can also be extracted from ordinary relational structures; the multidimensional query processor in this case translates the multidimensional queries into SQL queries that are executed by the relational DBMS.

Specific OLAP products, as a rule, are: a multidimensional data presentation tool, an OLAP client (for example, Pivot Tables in Microsoft Excel or ProClarity from Knosys), a multidimensional server DBMS, an OLAP server (for example, Oracle Express Server or Microsoft OLAP Services).

The multidimensional processing layer is usually embedded in an OLAP client and / or an OLAP server, but it can be selected in its pure form, such as the Microsoft Pivot Table Service component.

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch

electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Tasks

1. Install data mining add-ins for Microsoft Office 2007. Configure MS SQL Server to work with add-ins. Create and test the connection.

2. Format the prepared data set as a table.

Make sure you can access the tab with the predictive tables analysis tools.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>

2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>

3. <http://www.intuit.ru/studies/courses/2312/612/info>

4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / LP. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>

6. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 2. Using the AnalyzeKeyInfluencers and DetectCategories tools

Topic - Using AnalyzeKeyInfluencers tools and "DetectCategories"

The goal of the work is to acquire practical skills in using the tools "AnalyzeKeyInfluencers" and "DetectCategories"

Formed competences or their parts - DPC-2; PC-2; PC-3

The theoretical part.

Let's start the direct study of data mining tools. The package of add-ins for MS Office 2007 includes a spreadsheet with sample data. It can be opened from Start-> Data Mining Add-ins.

Microsoft SQL Server 2008.

The AnalyzeKeyInfluencers tool allows you to determine how the parameter you are interested in depends on others. At the same time, it is important to correctly determine what can depend on what. Actually, this is partly the skill of the analyst, based on his knowledge of the subject area and the DM methods used.

Due to the fact that we assess the degree of mutual influence of different parameters on each other, it is worth removing from the consideration completely independent at once and vice versa, completely dependent. Suppose, for example, we want to evaluate the influence of various factors on the level of a person's wages. If we have a field containing a unique identifier (for example, the sequence number of an entry in a table or a passport number), it should be removed from consideration, as it does not affect the value of the parameter under study. Another example is that we have a monthly and annual salary value calculated as monthly salary multiplied by 12. We know that these values are always related, looking for dependence of one on the other by means of DM does not make sense, and the existing strong dependence hide the influence of other factors that we just want to reveal.

After the start of the analysis procedure, a report on the factors of influence will be generated and an additional comparative report will be proposed.

If you add a comparative report for the two selected values, you can see the difference between the choice in favor of one value of the target column and the choice in favor of the other.

Task 1. Conduct the analysis in accordance with the considered example.

Task 2. On the same data set, analyze the dependence of income level on education, marital status, type of work, gender, age and region of the client's residence. Describe the results.

Task 3. Offer your version of data analysis, and an example of using the obtained results.

The Detect Categories tool allows you to solve the clustering problem, i.e. dividing the whole set of options into “natural” groups, the members of which are closest in terms of a number of features. This task is also called a segmentation task.

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Tasks

1. Rename the category Category 3.
2. Analyze the parameters that characterize the remaining categories, and give them meaningful names.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>
2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>
3. <http://www.intuit.ru/studies/courses/2312/612/info>
4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / LP. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>
6. <https://technet.microsoft.com>
7. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 3. Using the "FillFromExample" and "Forecast" tools

Topic - Using the "FillFromExample" and "Forecast" tools

The goal of the work is to acquire practical skills in using the tools "FillFromExample" and "Forecast"

Formed competences or their parts - GPC-2; PC-2; PC-3

The theoretical part.

Both considered tools are used to solve problems of predicting unknown values of parameters. Therefore, in both cases, a training data set is required, on the basis of which the model used for the prediction is built.

Filling in by example

The required data is on the "Filling from the example" sheet. It describes the number of customers of the store. For some of them, it is noted whether this client is highly profitable. These lines will be used as a training set. The task of the analysis will be the evaluation of other customers for this parameter.

To solve this problem, the Microsoft Logistic Regression algorithm is used. It should be understood that to create a model in a training set, variants with all possible values of the target column must be presented. The required number of examples depends on the characteristics of the subject area. But in many cases it is true that the more characteristic examples in the training sample, the more qualitatively the model will be trained.

Accordingly, this tool is unsuitable for the task of predicting the values of a parameter that can take continuous numerical values.

Another feature is that the analysis is carried out in columns (predicted column value). If the row that needs to be filled is stored as a string, before beginning the analysis, you need to transpose (copy to clipboard, select "Paste Special" in the context menu and check the "Transpose" box).

The task. Analyze and describe the results.

Forecast tool allows you to build a forecast of the values of the number series. The row should be represented by a column in the table (if the studied values are organized in the form of a string, it is required, as in the case of the tool "FillFromExample", to transpose).

The tool searches the analyzed sequence for the following types of templates:

- trend - the trend of changing values. A trend can be ascending (increasing values of a series) or descending (decreasing values);
- frequency (seasonality) - the event repeats itself through certain intervals;
- cross-correlation - dependence of the values of one series on others

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Tasks

1. Use the tool to build a sales forecast for the year (12 values). Analyze the schedule. In your opinion, what type of periodicity did the tool detect in the source data and use it for prediction?

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>

2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>

3. <http://www.intuit.ru/studies/courses/2312/612/info>

4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / L.P. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>

6. <https://technet.microsoft.com>

7. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 4. Using the "HighlightExceptions" and "ScenarioAnalysis" tools

Theme - Using the "HighlightExceptions" and "ScenarioAnalysis"

The purpose of the work - the acquisition of practical skills to use tools "HighlightExceptions" and "ScenarioAnalysis" "

Formed competences or their parts - GPC-2; PC-2; PC-3

The theoretical part.

As the name implies, the tool allows you to identify data that stands out among the existing set. This can be useful in some cases. First, it may be erroneous data (for example, the results of an operator error when entering some values). Secondly, exceptions may be of separate interest (as, for example, in the case of detection of fraudulent actions with bank cards, etc.). In addition, exception analysis can be considered as a preliminary part of data mining using other methods. In particular, this makes it possible to exclude non-typical examples from entering the training set.

During the work, the HighlightExceptions tool creates a temporary mining model using the MicrosoftClustering algorithm. For each analyzed line, the degree of belonging to the identified clusters is assessed. Values that are far away from all clusters are flagged as exceptions.

Task 1. Conduct an exception analysis for the Prediction dataset. Offer an interpretation of the results.

The Scenario Analysis tool allows you to model the effect of changing one of the parameters (values of one column) on another, associated with the first. The tool is based on the use of the Microsoft Logistic Regression algorithm. For the formation of a temporal model, an educational sample is required containing at least 50 records.

The Scenario Analysis tool includes two component parts - "Analysis of the search for solutions" (GoalSeek) and "Analysis of possible options" ("What-If").

Using the GoalSeek tool allows you to assess whether we can achieve the desired value in the target column by changing the values of the selected parameter. The tool allows you to analyze for one record, and for the entire table.

Using this tool you need to be prepared that not all the variants of the request can be answered. This may be due to the fact that there are no combinations of interest in the source data. There may also be problems due to data types.

In addition, we must not forget that the request must be formed taking into account knowledge of the subject area.

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock

frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Tasks

The task. Perform analysis for a single row and table, similar to that described above. Comment on the results.

The What-If tool allows you to solve the problem inverse to GoalSeek: evaluate the value of the target variable for a certain change in a given parameter.

The task. Perform data analysis

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>

2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>

3. <http://www.intuit.ru/studies/courses/2312/612/info>

4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / L.P. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>

6. <https://technet.microsoft.com>

7. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 5. Using the "Prediction Calculator" and "ShoppingbasketAnalysis" tools

Topic - Using the "Prediction Calculator" and "ShoppingbasketAnalysis" tools

The purpose of the work - the acquisition of practical skills to use tools "Prediction Calculator" and "ShoppingbasketAnalysis"

Formed competences or their parts - GPC-2; PC-2; PC-3

Theoretical part.

Prediction Calculator helps you generate and customize "calculator", which allows you to estimate the chances of getting the expected value of the target parameter without connecting to the analytical services of SQL Server. In particular, this feature can be very useful for remote users.

According to the results of the analysis, a forecast is formed, which can be assigned to one of four categories:

- true positive outlook (TruePositive) - correct forecast. For example, a client for whom the forecast showed the truth is actually interested in buying a bike. The store made a profit;
- True negative outlook (TrueNegative) - true negative outlook. The client, for whom the forecast showed disinterest in buying, is not really going to buy a bike. The store did not make a profit, but did not incur costs (for sending promotional offers, etc.);
- a false positive outlook (FalsePositive; type 1 error) is incorrect, indicating that the customer wants to make a purchase, although in fact this is not the case (the store can lead to customer escort costs);
- a false negative outlook (FalseNegative; error 2 kind) incorrect forecast, showing that the client does not want to make a purchase, although in fact he is interested in it (may lead to lost profits).

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this;

observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Tasks

The task. Perform analysis for two different sets of values of the profits from the true predictions and losses from false. Comment on the results.

Tool Shopping Basket Analysis. It allows, for example, on the basis of data on purchases to highlight the goods most often found in one order, and to formulate recommendations for joint sales.

Task 1. Conduct an analysis similar to that described above.

Task 2. Analyze which goods (and not the categories of goods, as it was before), are acquired together. Describe the results obtained.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>
2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>
3. <http://www.intuit.ru/studies/courses/2312/612/info>
4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>
5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / L.P. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>
6. <https://technet.microsoft.com>
7. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 6. Using Data Mining Client Tools for Excel 2007 to prepare data

Topic - Using Data Mining Client for Excel 2007 tools to prepare data

The goal of the work is to acquire practical skills in using Data Mining Client for Excel 2007 tools for data preparation

Formed competences or their parts - GPC-2; PC-2; PC-3

The theoretical part.

The "Table Analysis Tools for Excel" (TableAnalysisTools) discussed in previous labs allow you to quickly "standard" analysis of available data. At the same time, this toolkit does not provide specific opportunities for preparing data for analysis, evaluation of results, etc. From Excel, this can be done using a data mining client (DataMiningClient), which is also included in the set of data mining add-ons. During the "Data Mining Add-in for MicrosoftOffice", it was noted that it is desirable to make a complete installation of add-ins, which includes DataMiningClient.

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Tasks

The task. Spend described in the laboratory processing of the selected data set.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of

Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php? Page = book & id = 233071>

2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>

3. <http://www.intuit.ru/studies/courses/2312/612/info>

4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / LP. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php? Page = book & id = 118188>

6. <https://technet.microsoft.com>

7. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 7. Use Data Mining Client Tools for Excel 2007 to create a mining model

The goal of the work is to acquire practical skills in using Data Mining Client for Excel 2007 tools to create a mining model.

Formed competences or their parts - GPC-2; PC-2; PC-3

Theoretical part.

Data Modeling tools for creating data mining models:

(Classify)	creates a classification model based on existing Excel spreadsheet data, an Excel range, or an external data source (AnalysisServicesDataSource). On the basis of the data being processed, templates are formed which, when used, allow us to attribute the considered example to one of the possible classes. The default algorithm is DecisionTrees, but LogisticRegression, NaiveBayes, NeuralNetworks are also available.
(Estimate)	allows you to create a model for estimating the value of the target parameter (it must be numeric) based on data from a table or a range of Excel cells or an external data source. The default is Decision Trees, also available in Linear Regression, Logistic Regression, Neural Networks.
(Cluster)	launches a wizard that allows you to build a clustering model based on data from a table or Excel range, or an external data source. The model defines groups of lines with similar characteristics, for which the MicrosoftClustering algorithm is used. This task is similar to the solved DetectCategories tool from the TableAnalysisTools set.
(Associate)	helps to create a model that describes the relationship of objects (purchased goods, etc.) affected by a single transaction, for which the AssociationRules algorithm is used. We encountered a similar task using the ShoppingBasketAnalysis tool from TableAnalysisTools. To build an analysis model, it is necessary that the source data contain a column with a transaction identifier and be sorted by it. Only a spreadsheet or a range of Excel cells can be used as a data source.
(Forecast)	This wizard allows you to build a model to predict new values in numerical sequence, similar to the Forecast tool in TableAnalysisTools. The TimeSeries algorithm is used, for which it is required that the column (or columns) for which the forecast will be performed have continuous numerical values. There may also be a column with a time stamp (in this case, the rows in the table should be sorted by it).
(Advanced)	allows you to create a data mining structure or add a new model to an existing structure (for example, to compare the results of different analysis algorithms)

Use the Classify tool. In the dataset that comes with the add-ins (Start menu - > Data Mining Add-ons -> Excel Data Samples), select the courseData table containing a random sample of 70% of the data from the SourceData table. Launch the Classify wizard, in the first window of which there will be a comment on the use of the tool, and the second window will allow you to specify the data source for analysis (TrainingData table).

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Tasks

Task 1. Create a mining model similar to that described in the lab.

Task 2. Use the dataset in the Associate table and the wizard of the same name to create a model that describes the relationship between product categories in one order. If necessary, use the tool help system. Analyze the identified rules and dependency diagram.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>
2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>
3. <http://www.intuit.ru/studies/courses/2312/612/info>
4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

Lab number 8. Analysis of forecast accuracy and use of a mining model

The purpose of the work is to acquire practical skills in analyzing the accuracy of the forecast and using the mining model

Formed competences or their parts - GPC-2; PC-2; PC-3

Theoretical part.

In a previous lab, we created a model for classifying customers of a store in order to determine whether a customer will make a purchase or not. The next task is to assess the accuracy of the constructed mining model. To do this, you can use tools from the Accuracy and Validation group.

Task 1. Build a precision diagram similar to the one presented above (the file used is Excel Data Samples). Additionally, build a chart for BikeBuyer = "No". Explain the difference in the appearance of the graphs.

Task 2. In the previous task for testing purposes we used data from the model. The model was formed on the data from the table TrainingData. The TestingData table contains 30% of the data from the SourceData source set. Check the accuracy of the model on the TestingData set.

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Tasks

The task. Query the mining model. Rate the results.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of

Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>

2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>

3. <http://www.intuit.ru/studies/courses/2312/612/info>

4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / L.P. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>

6. <https://technet.microsoft.com>

7. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 9. Build a clustering model

The goal of the work is to acquire practical skills in building a clustering model.

Formed competences or their parts - GPC-2; PC-2; PC-3

The theoretical part.

There are various approaches to the classification of clusters, while using a fairly wide range of characteristics, according to which enterprises are combined into homogeneous groups (clusters). Most often, enterprise clusters are classified according to such parameters as: geographical location, availability of capital, proximity of suppliers, competitiveness, access to specialized services, development of labor potential, availability of specialized educational institutions and research organizations, industry affiliation, etc.

In the scientific literature there are various approaches to the principles of cluster formation. At present, seven main approaches are described that occur in practice and represent a combination of various factors on which, in turn, the choice of a particular cluster strategy is based. Let us briefly list these models:

1. The geographical model involves the construction of spatial clusters of various economic activity from purely local (for example, horticulture in the Netherlands) to truly global (aerospace cluster).

2. The horizontal model is used to describe those cases where several industries are included in a larger cluster.

3. Vertical model: enterprises in adjacent stages of the production process may be present in clusters. At the same time, it is important who exactly of the network participants is the initiator and ultimate performer of innovations within the cluster.

4. Literal model: a cluster combines different sectors that can provide economies of scale, leading to new combinations (for example, a multimedia cluster).

5. Technological model: describes the set of enterprises in various industries using the same technology (for example, a biotechnological cluster).

6. Focus model: a cluster of firms centered around a single center — an enterprise, a research institute or a training center, and who are consumers of knowledge and innovation [3].

7. Qualitative model: characterizes in quality categories the availability and methods of interaction of firms within the cluster. For her, it is not only essential whether firms do cooperate, but also how they do it [2].

Clusters can vary in size, breadth of coverage, level of development, depending on industries. The nature of industries allows changing the boundaries of clusters due to the emergence of new industries or firms, as well as as a result of the development or changes in the business environment. It should be noted that clusters

are found within the framework of both developed and developing countries. The difference will be the presence in specialized industrial clusters of specialized supplier bases, the most extensive array of related industries, and deeper connections with consumers [1].

Currently, in the formation of industrial clusters should be based on a set of methods. The main ones are:

1. Analytical method: the study of data on the territory, its natural, labor and financial resources of the productive forces.
2. The method of indicative planning of competitiveness: quantitative and qualitative analyzes that exist and the definition of prospective competitiveness.
3. Method of modeling: the study of cluster objects using the construction and study of models of real-life organizations, processes or phenomena.
4. The input-output method: a study of the circulation of goods and services between industries and other sectors of the economy.

In practice, the formation of industrial clusters, these methods are used in combination, as taken separately they do not reflect an objective economic situation. The analytical method displays the situation on the market at a specific point in time (data on enterprises, their productivity, number of employees, etc.). The method of indicative planning consists in determining economic priorities. Indicative plan is a recommendation. The main advantages of indicative planning are reduced uncertainty in decision making. The simulation method is a method that allows you to build models that describe the processes as they would in reality. This method allows you to avoid any errors in the future. The “expenditure issue” method examines the data on the circulation of goods based on a set of economic indicators (production capacity, productivity growth, output, etc.).

Also practice distinguishes several stages of creating and managing clusters.

1st stage: identification of clusters (methods of their selection). There are many paired categories to characterize the process and cluster identification techniques, but most of them come from two main approaches. In the first, which can be called "bottom", clusters are identified on a specifically selected territory, based on the presence of previously known enterprises and leading industries. The second approach uses a technique, conventionally called "top", where spatial localizations of production are sought, oriented to specific types of economic activity. Approaches to identifying clusters “from above,” taking into account two invariable characteristics of clusters — functional connectivity and geographical proximity — are traditionally divided into two types:

- functional, focused on the identification of industrial clusters;
- spatial, geographic clusters oriented.

Stage 2: Manage existing clusters.

3rd stage: Evaluation of the economic effect of creating a cluster.

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>

2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>

3. <http://www.intuit.ru/studies/courses/2312/612/info>

4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / L.P. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>

6. <https://technet.microsoft.com>

7. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 10. Creating a data source view

The purpose of the work - the acquisition of practical skills to create data source views

Formed competences or their parts - GPC-2; PC-2; PC-3

The theoretical part.

The data sources view contains a logical model of the schema used by the Analysis Services multidimensional database objects, namely cubes, dimensions, and mining structures. The data source view is a metadata definition stored in XML format for schema elements used in the Unified Multidimensional Model (UDM) and Data Mining Structures. Data source view:

- contains metadata representing selected objects from one or more basic data sources, or metadata that will be used to form a basic relational data repository using a top-down approach in building a schema.
- Can be built on the basis of one or several data sources, which allows you to define multidimensional objects and data mining objects that combine data from several sources.
- may contain relationships, primary keys, object names, calculated columns, and queries that are missing from the core data source and that exist separately from it;
- not visible, but available for requests from client applications.

The data source view is a necessary component of the multidimensional model. Most Analysis Services developers create a data source view in the early stages of model design, providing at least one data source view based on an external relational database that provides basic data. However, you can also create a data source view in a later phase by creating a schema and basic database structure after creating dimensions and cubes. This second approach is sometimes called descending and is often used for prototyping and analytical modeling. In this case, use the schema creation wizard to create a basic view of data sources and data source objects based on OLAP objects defined in an Analysis Services project or database. Regardless of how and when a data source view is created, it must be in each model before it can be processed.

The data source view contains the following elements.

- Name and description.
- Identification of any subset of the scheme, obtained from one or more data sources, up to the full scheme and including it, including:
 - o table names; o column names; o data types; o valid null values; o the length of the columns; o Primary keys.
 - o communication "primary-foreign key".

- Notes to the schema from basic data sources, including the following:
 - o clear names of tables, views, and columns;
 - o named queries that return columns from one or more data sources (which appear as tables in the schema);
 - o Named calculations returning columns from a data source (displayed as columns of tables or views);
 - o logical primary keys (necessary if the primary key in the base table is not defined or not included in the view or named query);
 - o a logical primary key-foreign key relationship between tables, views, and named queries.

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>
2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>
3. <http://www.intuit.ru/studies/courses/2312/612/info>

4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>

5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / LP. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>

6. <https://technet.microsoft.com>

7. <http://www.intuit.ru/studies/courses/2312/612/info>

Lab number 11. Creating the structure and model of mining. Clustering task

The goal of the work is to acquire practical skills in creating the structure and model of intellectual analysis.

Formed competences or their parts - GPC-2; PC-2; PC-3

The theoretical part.

The task of cluster analysis is to divide the set of objects G into m (m - integer) clusters (subsets) $Q_1, Q_2, \dots, Q_m$ based on the data contained in set X , so that each object G_j belongs to one and only one subset of the partition and that the objects belonging to the same cluster were similar, while the objects belonging to different clusters were heterogeneous.

The first step in creating a sequence clustering data mining model is to use a data mining wizard to create a new data mining structure and data mining model based on the sequence clustering algorithm.

For this purpose, the data source view will be used, which was used to analyze consumer behavior, to which is added a column containing a sequence identifier. In this scenario, "sequence" means the order in which customers add items to their shopping cart during purchases.

There will also be added a few columns that are used in one of the models to group buyers by demographic. Creating a sequence clustering structure and model

1. In Solution Explorer in SQL Server Data Tools (SSDT), right-click the Mining Structures item and select Create Mining Structure.

2. On the Welcome to the Data Mining Wizard page, click Next.

3. On the Select Definition Method page, make sure that Based on an existing relational database or data store is selected, and then click Next.

4. On the Create Mining Structure page, verify that the Create Mining Structure and Mining Model option was selected. Then click the drop-down list. Which data mining method to use? and select Microsoft Sequence Clustering Algorithm. Press "Next."

The Select Data Source View page opens. Under Available data source views, select Orders.

Orders is the data source view that was used for the shopping cart scenario.

5. Click Next.

6. On the Definition of Table Types page, select the Option check box next to the vAssocSeqOrders table and the Attachment check box next to the vAssocSeqLineItems table. Press "Next."

7. On the Definition of Training Data page, select the columns to be used in the model by checking the box as follows.

- o IncomeGroup Select the Enter checkbox.

This column contains useful customer information that can be used for clustering. Such information will be used in the first model and will not be taken into account in the second.

- o OrderNumber Select the Key checkbox.

This field will be used as the identifier of the table of options or the Key value. In general, the key field of the variation table is never used as input, because the key contains unique values that are not of interest for clustering.

- o Region Select the Enter checkbox.

This column contains useful customer information that can be used for clustering. Such information will be used in the first model and will not be taken into account in the second.

- o LineNumber Select the Key and Input checkboxes.

The LineNumber field will be used as an identifier for the nested table or the Sequence Key value. The key for a nested table should always be used for input.

- o Model Select the Enter and Predicted checkboxes.

Make sure that all checkboxes are selected correctly, then click Next.

8. On the Definition of Content and Data Type of Columns page, verify that the grid contains the columns, content types, and data types listed in the following table, and then click Next.

Equipment and materials

To perform laboratory work, it is recommended to use a personal computer with the following characteristics: 32-bit (x86) or 64-bit (x64) processor with a clock frequency of 1 GHz and above, RAM - 1 GB and above, free disk space - at least 1 GB graphics device DirectX 9. Software:

operating system WINDOWS 7 and higher, SQL Server, MS Excel.

Safety instructions

Safety in the performance of laboratory work coincides with the generally accepted for users of personal computers. Independently do not repair the personal computer, install and uninstall software; in case of malfunction of the personal computer, inform the laboratory personnel (operator, administrator) about this; observe safety regulations when working with electrical equipment; do not touch electrical outlets with metal objects; the workplace of the user of the personal computer must be kept clean; It is not allowed to take food or drinks near the personal computer.

Report content and form

A brief summary of the theoretical justification;

Description of the main stages of the work.

References recommended for use on this topic:

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>
2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>
3. <http://www.intuit.ru/studies/courses/2312/612/info>
4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>
5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / L.P. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>

RECOMMENDED LITERATURE

1. Grekul, V.I. Designing information systems [Electronic resource] / V.I. Grekul, G.N. Denischenko, N.L. Korovkina. - Moscow: Internet University of Information Technologies, 2005. - 304 p. - ISBN 59556-0033-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=233071>
2. Tereshchenko, P.V. Interfaces of information systems: a tutorial [Electronic resource] / P.V. Tereshchenko, V.A. Astapchuk. - Novosibirsk: NSTU, 2012. - 67 p. - ISBN 978-5-7782-2036-2. - URL: <http://biblioclub.ru/index.php?page=book&id=228775>
3. <http://www.intuit.ru/studies/courses/2312/612/info>
4. Malyshev, S.L. Basics of the Internet economy: a tutorial [Electronic resource] / S.L. Malyshev. - M.: Eurasian Open Institute, 2011. - 120 p. - ISBN 978-5-374-00556-1. - URL: <http://biblioclub.ru/index.php?page=book&id=90789>
5. Gavrilov, L.P. Basics of e-commerce and business [Electronic resource] / L.P. Gavrilov. - M.: SOLON-PRESS, 2009. - 592 p. - ISBN 978-5-91359-065-7. - URL: <http://biblioclub.ru/index.php?Page=book&id=118188>

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN
FEDERATION

Federal State Autonomous Educational Institution for Higher Education
«NORTH-CAUCASUS FEDERAL UNIVERSITY»

METHODICAL INSTRUCTIONS

on the organization of independent work on the discipline " Multidimensional data
analysis " for the areas of training 09.04.03 "Applied Information Science"
(Directivity (profile) "Knowledge Management")

Stavropol, 2019

Introduction

The discipline "Multidimensional Data Analysis" forms professional skills of a graduate student in the direction of training 09.04.03 "Applied Information Science".

The study of the discipline "Multidimensional Data Analysis" is based on laboratory work and individual lessons under the guidance of a teacher, independent work of students. The objectives of the course are:

1. the formation of the master's personal and professional qualities, allowing to carry out professional activities related to analysis, development and implementation of information and analytical systems;
2. The study by the masters of the problems and areas of the use of methods for automating the analysis of information preparation for making managerial decisions with the use of modern wide-use toolkits and specialized application packages;
3. knowledge of the fundamentals of the problems and areas of the use of artificial intelligence, expert and knowledge-based systems.

1. General characteristics of the independent work of the student in the study of the discipline Multidimensional data analysis - the formation of a set of professional competencies of the future master in the direction of training 09.04.03 "Applied Information Science".

2. Tasks of mastering the discipline:

1. the formation of the master's personal and professional qualities, allowing to carry out professional activities related to analysis, development and implementation of information and analytical systems;

2. The study by the masters of the problems and areas of the use of methods for automating the analysis of information preparation for making managerial decisions with the use of modern wide-use toolkits and specialized application packages;

3. knowledge of the fundamentals of the problems and areas of the use of artificial intelligence, expert and knowledge-based systems.

3. Formed competencies - GPC-2; PC-2; PC-3

4. Schedule of independent work

No	Discipline section	Semester week	Independent work of the student in hours	Forms of monitoring progress
1	Lab number 1. OLAP application architecture	1.	6	Interview on the topic
2	Lab number 2. Using the AnalyzeKeyInfluencers and DetectCategories tools	2.	6	Interview on the topic
3	Lab number 3. Using the "FillFromExample" and "Forecast" tools	3.	6	Interview on the topic
4	Lab number 4. Using the "HighlightExceptions" and "ScenarioAnalysis" tools	4.	6	Interview on the topic
5	Lab number 5. Using the "Prediction Calculator" and "ShoppingbasketAnalysis" tools	5.	6	Interview on the topic
6	Lab number 6. Using Data Mining Client Tools for Excel 2007 to prepare data	6.	6	Interview on the topic

7	Lab number 7. Use Data Mining Client Tools for Excel 2007 to create a mining model	7.	7,5	Interview on the topic
8	Lab number 8. Analysis of forecast accuracy and use of a mining model	8.	6	Interview on the topic
9	Lab number 9. Build a clustering model	9.	6	Interview on the topic
10	Lab number 10. Creating a data source view	10.	6	Interview on the topic
11	Lab number 11. Creating the structure and model of mining. Clustering task	11.	6	Interview on the topic
TOTAL			67,5	

5. Methodical recommendations for the study of theoretical material

At the first stage, it is necessary to familiarize yourself with the work program of the discipline, which examines the content of the topics of the discipline of the lecture course, the relationship of the topics of lectures with practical exercises, topics and types of independent work. For each type of independent work, certain reporting forms are provided.

Questions for the protection of tasks for laboratory work

Lab number 1. OLAP application architecture

1. Basic data mining add-ins for MS Office 2007
2. Data Mining Add-in for MSSQLServer

Lab number 2. Using the AnalyzeKeyInfluencers and DetectCategories tools

1. Data mining add-ins
2. Analysis of key factors of influence
3. Category Detection

Lab number 3. Using the "FillFromExample" and "Forecast" tools

1. Using the trend pattern type sequence is the trend of changing values. The trend can be upward (increase in the values of the series) or downward (decrease in values)
2. The use of a pattern sequence like periodicity (seasonality) - the event repeats itself at certain intervals;

3. The use of the pattern sequence of the type of cross-correlation - the dependence of the values of one series from the others

Lab number 4. Using the "HighlightExceptions" and "ScenarioAnalysis" tools

1. Highlighting Exceptions
2. Scenario analysis
3. Analysis of possible options

Lab number 5. Using the "Prediction Calculator" and "ShoppingbasketAnalysis" tools

1. Calculation of the forecast
2. Categories of the forecast
3. Analysis of the shopping basket

Lab number 6. Using Data Mining Client Tools for Excel 2007 to prepare data

1. CleanData tool

Lab number 7. Use Data Mining Client Tools for Excel 2007 to create a mining model

1. Create a mining model.

Lab number 8. Analysis of forecast accuracy and use of a mining model

1. Chart accuracy.
2. Classification matrix
3. Decision Tree
4. Requests to the DM model

Lab number 9. Build a clustering model

1. DMX structure processing request

Lab number 10. Creating a data source view

1. Work with data source
2. Application Deployment Modes

Lab number 11. Creating the structure and model of mining. Clustering task

1. The concept of representation

2. Creating a view
3. Creating Named Calculations and Queries

6. Guidelines for the preparation for offset

The procedure of credit as a separate control measure is not carried out, the assessment of the student's knowledge is based on the results of current control.

7. Methodological materials defining the procedures for evaluating knowledge, skills, abilities and (or) experience, characterizing the stages of the formation of competencies

The current attestation of graduate students is conducted by teachers conducting practical classes in the discipline, based on the results of an oral survey on practical classes and the performance of independent work tasks.

Undergraduate student should prepare answers to the interview questions for self-study, perform additional tasks on the topic of the lesson, work out additional scientific sources, make an outline of the article (if provided). The student receives the maximum mark if he actively participates in the work, owns the material, is able to logically and clearly express his thoughts, creatively approaches the solution of the main issues of the topic, shows the independence of thinking.

The basis for the lower scores are:

- poor knowledge of the topic and basic terminology;
- lack of ability to apply theoretical knowledge to solve practical problems.