Министерство науки и высшего образования Российской Федерации

федеральное государственное автономное образовательное учреждение высшего образования

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»

(Университет ИТМО)

Факультет программной инженерии и компьютерной техники

Отчет по лабораторной работе №4

по дисциплине « База данных »

Тема: Data Mining на основе модели CRIPS-DM

Выполнила:

студентка гр.№P33212 Ян Цзяфэн

Санкт-Петербург

2020

Оглавление

Задание

Для выполнения лабораторной работы №4 необходимо:

- Реализовать разработанную в рамках лабораторной работы №3 даталогическую модель в реляционной СУБД PostgreSQL.
- Заполнить созданные таблицы данными.
- Обеспечить целостность данных при помощи средств языка DDL.
- В рамках лабораторной работы должны быть разработаны скрипты для создания/удаления требуемых объектов базы данных, заполнения/удаления содержимого созданных таблиц.

Отчёт по лабораторной работе должен содержать:

- титульный лист;
- текст задания;
- описание предметной области;
- DDL-скрипты, часть DML-скриптов;
- выводы по работе;

Темы для подготовки к защите лабораторной работы:

1. Язык DDL
2. Обеспечение целостности данных
3. Язык DML

Описание предметной области

В этой предметной области будет анализироваться интеллектуальный анализ данных на основе модели CRIPS-DM. Интеллектуальный анализ данных (Data Mining), также известный как обнаружение знаний (Knowledge Discover in Database, KDD) в базах данных, в настоящее время является актуальной проблемой в области искусственного интеллекта и исследований баз данных. В проекте CRISP-DM (CRoss Industry Standard Process for Data Mining) была предложена комплексная модель процесса для выполнения проектов интеллектуального анализа данных. Модель процесса не зависит

ни от отрасли, ни от используемой технологии.

Предметная область описывает типичные проблемы интеллектуального анализа данных, которые решаются с помощью процесса. У каждого процесса есть соответствующие подпроцессы, также называемые задачами. Выполнение каждой задачи требует от пользователя выполнения определенных действий для получения выходных данных. Результаты задачи могут быть сведены в несколько частей отчета. Технологии, необходимые для каждого типа проблем, различаются, и алгоритмы, которые необходимые для реализации технологий, также различны. Кроме того, данная база данных также включает связанные функции и атрибуты набора данных, которые можно использовать.

Даталогическая модель

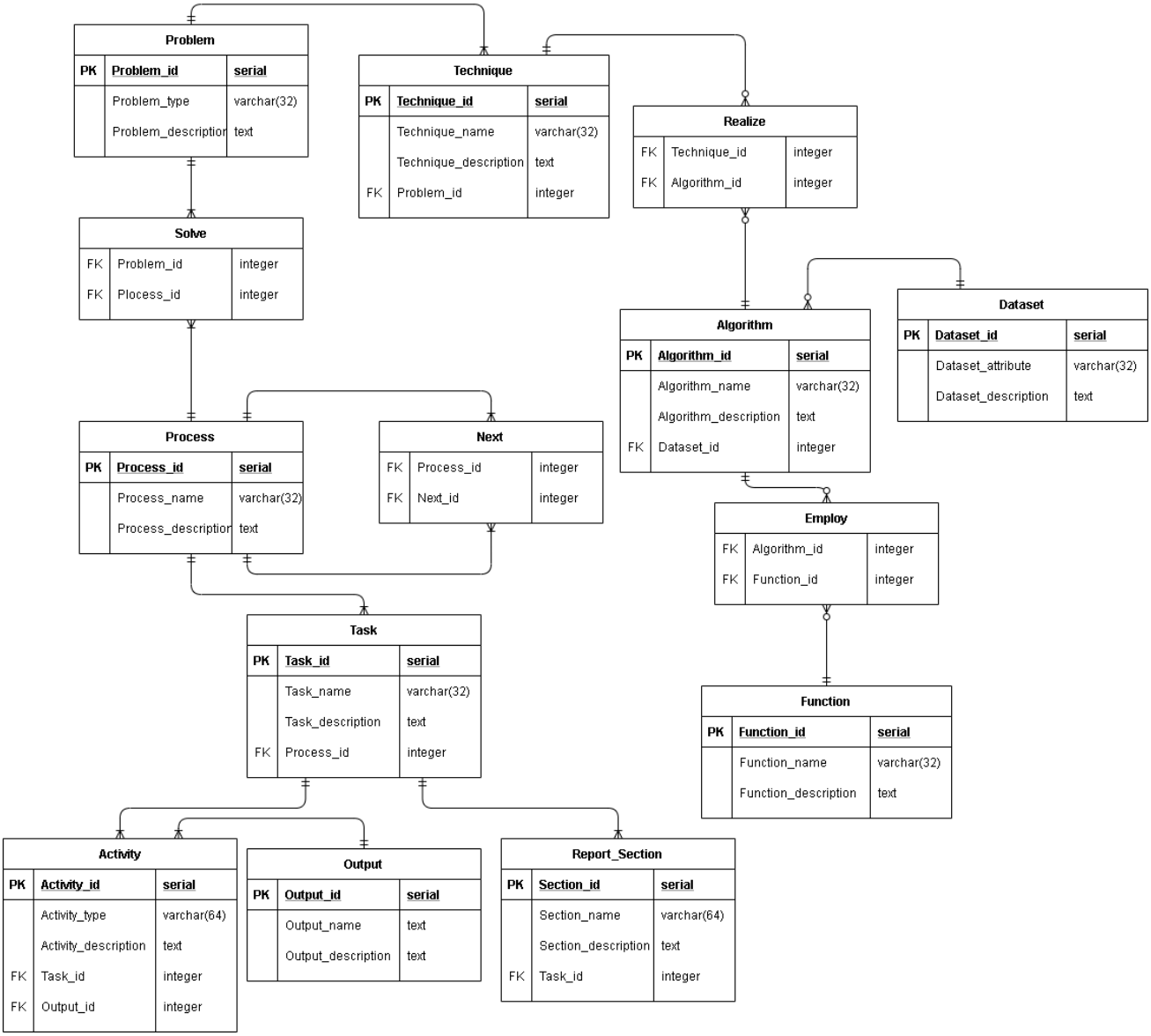

Рис.1 даталогическая модель

DDL-скрипты

Скрипты для создания требуемых объектов базы данных

```
CREATE TABLE Problem (
   Problem_id serial PRIMARY KEY,
   Problem_type varchar (64),
   Problem_description text NOT NULL
);
CREATE TABLE Process (
   Process_id serial PRIMARY KEY,
   Process_name varchar (32) NOT NULL,
   Process_description text NOT NULL
);
CREATE TABLE Solve (
   Problem_id integer REFERENCES Problem,
   Process_id integer REFERENCES Process,
   UNIQUE (Problem_id, Process_id)
);
CREATE TABLE Next (
   Process_id integer REFERENCES Process,
   Next_id integer REFERENCES Process (Process_id),
   UNIQUE (Process_id, Next_id),
   CHECK (Process_id != Next_id)
);
CREATE TABLE Task (
   Task_id serial PRIMARY KEY,
   Task_name varchar (32) NOT NULL,
   Task_description text NOT NULL,
   Process_id integer REFERENCES Process
);
CREATE TABLE Output (
   Output_id serial PRIMARY KEY,
   Output_name text NOT NULL,
   Output_description text NOT NULL
);
CREATE TABLE Activity (
   Activity_id serial PRIMARY KEY,
   Activity_type varchar (64),
   Activity_description text NOT NULL,
   Task_id integer REFERENCES Task,
   Output_id integer REFERENCES Output
);
```

```
CREATE TABLE Report_Section (
   Section_id serial PRIMARY KEY,
   Section_name varchar (64) NOT NULL,
   Section_description text NOT NULL,
   Task_id integer REFERENCES Task
);
CREATE TABLE Technique (
   Technique_id serial PRIMARY KEY,
   Technique_name varchar (32) NOT NULL,
   Technique_description text NOT NULL,
   Problem_id integer REFERENCES Problem
);
CREATE TABLE Dataset (
   Dateset_id serial PRIMARY KEY,
   Dataset_attribute varchar (32) NOT NULL,
   Dataset_description text
);
CREATE TABLE Algorithm (
   Algorithm_id serial PRIMARY KEY,
   Algorithm_name varchar (32) NOT NULL,
   Algorithm_description text NOT NULL,
   Dataset_id integer REFERENCES Dataset
);
CREATE TABLE Realize (
   Technique_id integer REFERENCES Technique,
   Algorithm_id integer REFERENCES Algorithm,
   UNIQUE (Technique_id, Algorithm_id)
);
CREATE TABLE Function (
   Function_id serial PRIMARY KEY,
   Function_name varchar (32) NOT NULL,
   Function_description text
);
CREATE TABLE Employ (
   Algorithm_id integer REFERENCES Algorithm,
   Function_id integer REFERENCES Function,
   UNIQUE (Algorithm_id, Function_id)
);
```

Скрипты для удаления требуемых объектов базы данных

DROP TABLE Employ;
DROP TABLE Function;
DROP TABLE Realize;
DROP TABLE Algorithm;
DROP TABLE Dataset;
DROP TABLE Technique;
DROP TABLE Report_Section;
DROP TABLE Activity;
DROP TABLE Output;
DROP TABLE Task;
DROP TABLE Next;
DROP TABLE Solve;
DROP TABLE Process;
DROP TABLE Problem;

DML-скрипты

Заполнения

INSERT INTO Problem(problem_type, problem_description) **VALUES** ('Data description and summarization', 'Data description and summarization aims at the concise description of characteristics of the data, typically in elementary and aggregated form. This gives the user an overview of the structure of the data.');

INSERT INTO Problem(problem_type, problem_description) **VALUES** ('Segmentation', 'Segmentation aims at the separation of the data into interesting and meaningful subgroups or classes. All members of a subgroup share common characteristics.');

INSERT INTO Problem(problem_type, problem_description) **VALUES** ('Concept descriptions', 'Concept description aims at an understandable description of concepts or classes. The purpose is not to develop complete models with high prediction accuracy, but to gain insights.');

	problem_id	problem_description	problem_type
1	1	Data description and summa…	Data description and summa…
2	2	Segmentation aims at the s…	Segmentation
3	3	Concept description aims a…	Concept descriptions
4	4	Classification assumes tha…	Classification
5	5	The aim of prediction is t…	Prediction
6	6	Dependency analysis consis…	Dependency analysis

INSERT INTO Process(process_name, process_description) **VALUES** ('Business understanding', 'This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.');

INSERT INTO Process(process_name, process_description) **VALUES** ('Data understanding', 'The data understanding phase starts with initial data collection and proceeds with activities that enable you to become familiar with the data, identify data quality problems, discover first insights into the data, and/or detect interesting subsets to form hypotheses regarding hidden information.');

INSERT INTO Process(process_name, process_description) **VALUES** ('Data preparation', 'The data preparation phase covers all activities needed to construct the final dataset [data that will be fed into the modeling tool(s)] from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record, and attribute selection, as well as transformation and cleaning of data for modeling tools.');

INSERT INTO Process(process_name, process_description) **VALUES** ('Modeling', 'In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, going back to the data preparation phase is often necessary.');

	process_id	process_name	process_description
1	1	Business understanding	This initial phase focuses on understanding the…
2	2	Data understanding	The data understanding phase starts with initia…
3	3	Data preparation	The data preparation phase covers all activitie…
4	4	Modeling	In this phase, various modeling techniques are …
5	5	Evaluation	At this stage in the project, you have built a …
6	6	Deployment	Creation of the model is generally not the end …

INSERT INTO Solve(PROBLEM_ID, PROCESS_ID) **VALUES** (1,1);
INSERT INTO Solve(PROBLEM_ID, PROCESS_ID) **VALUES** (1,2);
INSERT INTO Solve(PROBLEM_ID, PROCESS_ID) **VALUES** (1,3);
INSERT INTO Solve(PROBLEM_ID, PROCESS_ID) **VALUES** (1,4);
INSERT INTO Solve(PROBLEM_ID, PROCESS_ID) **VALUES** (1,5);
INSERT INTO Solve(PROBLEM_ID, PROCESS_ID) **VALUES** (1,6);

	problem_id	process_id
1	1	1
2	1	2
3	1	3
4	1	4
5	1	5
6	1	6

INSERT INTO Next(PROCESS_ID, NEXT_ID) **VALUES** (1, 2);
INSERT INTO Next(PROCESS_ID, NEXT_ID) **VALUES** (2, 3);
INSERT INTO Next(PROCESS_ID, NEXT_ID) **VALUES** (2, 1);
INSERT INTO Next(PROCESS_ID, NEXT_ID) **VALUES** (3, 4);
INSERT INTO Next(PROCESS_ID, NEXT_ID) **VALUES** (4, 5);
INSERT INTO Next(PROCESS_ID, NEXT_ID) **VALUES** (4, 3);
INSERT INTO Next(PROCESS_ID, NEXT_ID) **VALUES** (5, 6);
INSERT INTO Next(PROCESS_ID, NEXT_ID) **VALUES** (5, 1);

	process_id	next_id
1	1	2
2	2	3
3	2	1
4	3	4
5	4	5
6	4	3
7	5	6
8	5	1

INSERT INTO Task(TASK_DESCRIPTION, PROCESS_ID, TASK_NAME) **VALUES** ('The first objective of the data analyst is to thoroughly understand, from a business perspective, what the customer really wants to accomplish. Often the customer has many competing objectives and constraints that must be properly balanced. The analyst's goal is to uncover important factors, at the beginning, that can influence the outcome of the project. A possible consequence of neglecting this step is to expend a great deal of effort producing the right answers to the wrong questions.', 1, ' Determine business objectives');
INSERT INTO Task(TASK_DESCRIPTION, PROCESS_ID, TASK_NAME) **VALUES** ('This task involves more detailed fact-finding about all of the resources, constraints, assumptions, and other factors that should be considered in determining the data analysis goal and project plan. ', 1, 'Assess situation ');
INSERT INTO Task(TASK_DESCRIPTION, PROCESS_ID, TASK_NAME) VALUES ('A business goal states objectives in business terminology. A data mining goal states project objectives in technical terms.', 1, 'Determine data mining goals');
INSERT INTO Task(TASK_DESCRIPTION, PROCESS_ID, TASK_NAME) **VALUES** ('Describe the intended plan for achieving the data mining goals and thereby achieving the business goals. The plan should specify the steps to be performed during the rest of the project, including the initial selection of tools and techniques.', 1, 'Produce project plan');
INSERT INTO Task(TASK_DESCRIPTION, PROCESS_ID, TASK_NAME) **VALUES** ('Acquire the data (or access to the data) listed in the project resources. This initial collection includes data loading, if necessary for data understanding.', 2, 'Collect initial data ');
INSERT INTO Task(TASK_DESCRIPTION, PROCESS_ID, TASK_NAME) **VALUES** ('Examine the "gross" or "surface" properties of the acquired data and report on the results.', 2, 'Describe data');
INSERT INTO Task(TASK_DESCRIPTION, PROCESS_ID, TASK_NAME) **VALUES** ('This task addresses data mining questions using querying, visualization, and reporting techniques. These include distribution of key attributes relationships between pairs or small numbers of attributes, results of simple aggregations, properties of significant sub-populations, and simple statistical analyses. These

analyses may directly address the data mining goals; they may also contribute to or refine the data description and quality reports, and feed into the transformation and other data preparation steps needed for further analysis.', 2, 'Explore data ');
INSERT INTO Task(TASK_DESCRIPTION, PROCESS_ID, TASK_NAME) **VALUES** ('Examine the quality of the data, addressing questions such as: Is the data complete (does it cover all the cases required)? Is it correct, or does it contain errors and, if there are errors, how common are they? Are there missing values in the data? If so, how are they represented, where do they occur, and how common are they?', 2, 'Verify data quality');

task_id	task_description	process_id	task_name
1	The first objective of the data analys…	1	Determine business objectives
2	This task involves more detailed fact-…	1	Assess situation
3	A business goal states objectives in b…	1	Determine data mining goals
4	Describe the intended plan for achievi…	1	Produce project plan
5	Acquire the data (or access to the dat…	2	Collect initial data
6	Examine the “gross” or “surface” prope…	2	Describe data
7	This task addresses data mining questi…	2	Explore data
8	Examine the quality of the data, addre…	2	Verify data quality
9	Decide on the data to be used for anal…	3	Select data
10	Raise the data quality to the level re…	3	Clean data
11	This task includes constructive data p…	3	Construct data
12	These are methods whereby information …	3	Integrate data
13	Formatting transformations refer to pr…	3	Format data

INSERT INTO Output(OUTPUT_NAME, OUTPUT_DESCRIPTION) **VALUES** ('Background', 'Collate the information that is known about the organization’s business situation at the start of the project. These details not only serve to more closely identify the business goals to be achieved but also serve to identify resources, both human and material, that may be used or needed during the course of the project.');
INSERT INTO Output(OUTPUT_NAME, OUTPUT_DESCRIPTION) **VALUES** ('Business objectives', 'Describe the customer’s primary objective, from a business perspective. In addition to the primary business objective, there are typically a large number of related business questions that the customer would like to address.');
INSERT INTO Output(OUTPUT_NAME, OUTPUT_DESCRIPTION) **VALUES** ('Business success criteria', 'Describe the criteria for a successful or useful outcome to the project from the business point of view. This might be quite specific and readily measurable, or general and subjective. In the latter case, be sure to indicate who would make the subjective judgment.');
INSERT INTO Output(OUTPUT_NAME, OUTPUT_DESCRIPTION) **VALUES** ('Inventory of resources', 'List the resources available to the project, including personnel (business and data experts, technical support, data mining experts), data (fixed extracts, access to live warehoused or operational data), computing resources (hardware platforms), and software (data mining tools, other relevant software).');
INSERT INTO Output(OUTPUT_NAME, OUTPUT_DESCRIPTION) **VALUES** ('Requirements, assumptions, and constraints', 'List all requirements of the project, including schedule of completion, comprehensibility, and quality of results and security, as well as legal issues. List the assumptions made by the project. List the constraints made on the project.');
INSERT INTO Output(OUTPUT_NAME, OUTPUT_DESCRIPTION) **VALUES** ('Risks and contingencies',

'List the risks, that is, the events that might occur, impacting schedule, cost, or result. List the corresponding contingency plans: what action will be taken to avoid or minimize the impact or recover from the occurrence of the foreseen risks.');
INSERT INTO Output(OUTPUT_NAME, OUTPUT_DESCRIPTION) **VALUES** ('Terminology', 'Compile a glossary of terminology relevant to the project. This should include at least two components: (1) A glossary of relevant business terminology, which forms part of the business understanding available to the project (2) A glossary of data mining terminology, illustrated with examples relevant to the business problem in question');

output_id	output_description	output_name
1	Collate the information that is known about the organiza…	Background
2	Describe the customer's primary objective, from a busine…	Business objectives
3	Describe the criteria for a successful or useful outcome…	Business success criteria
4	List the resources available to the project, including p…	Inventory of resources
5	List all requirements of the project, including schedule…	Requirements, assumptions, and constraints
6	List the risks, that is, the events that might occur, im…	Risks and contingencies
7	Compile a glossary of terminology relevant to the projec…	Terminology
8	Prepare a cost-benefit analysis for the project, compari…	Costs and benefits
9	Describe the intended outputs of the project that enable…	Data mining goals
10	Define the criteria for a successful outcome to the proj…	Data mining success criteria
11	List the stages to be executed in the project, together …	Project plan
12	At the end of the first phase, the project team performs…	Initial assessment of tools and techniques
13	Describe all the various data used for the project, and …	Initial data collection report

INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Organization', 'Develop organizational charts identifying divisions, departments, and project groups. The chart should also identify managers' names and responsibilities', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Organization', 'Identify key persons in the business and their roles', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Organization', 'Identify an internal sponsor (financial sponsor and primary user/domain expert)', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Organization', 'Indicate if there is a steering committee and list members', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Organization', 'Identify the business units which are affected by the data mining project', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Problem area', 'Identify the problem area', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Problem area', 'Describe the problem in general terms', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Problem area', 'Check the current status of the project', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Problem area', 'Clarify prerequisites of the project', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Problem area', 'If necessary, prepare presentations and present data mining to the business', 1, 1);
INSERT INTO Activity(ACTIVITY_TYPE, ACTIVITY_DESCRIPTION, TASK_ID, OUTPUT_ID) **VALUES** ('Problem area', 'Identify target groups for the project result', 1, 1);

activity_id	activity_type	activity_description	task_id	output_id
1	Organization	Develop organizational cha…	1	1
2	Organization	Identify key persons in th…	1	1
3	Organization	Identify an internal spons…	1	1
4	Organization	Indicate if there is a ste…	1	1
5	Organization	Identify the business unit…	1	1
6	Problem area	Identify the problem area	1	1
7	Problem area	Describe the problem in ge…	1	1
8	Problem area	Check the current status o…	1	1
9	Problem area	Clarify prerequisites of t…	1	1
10	Problem area	If necessary, prepare pres…	1	1
11	Problem area	Identify target groups for…	1	1
12	Problem area	Identify the users' needs …	1	1
13	Current solution	Describe any solution curr…	1	1
14	Current solution	Describe the advantages an…	1	1
15	<null>	Informally describe the pr…	1	2
16	<null>	Specify all business quest…	1	2
17	<null>	Specify any other business…	1	2
18	<null>	Specify expected benefits …	1	2
19	<null>	Specify business success c…	1	3
20	<null>	Identify who assesses the …	1	3
21	Hardware resources	Identify the base hardware	2	4

INSERT INTO Report_Section(SECTION_NAME, SECTION_DESCRIPTION, TASK_ID) **VALUES** ('Background', 'The Background section provides a basic overview of the project context. This lists what area the project is working in, what problems have been identified, and why data mining appears to provide a solution. ', 1);
INSERT INTO Report_Section(SECTION_NAME, SECTION_DESCRIPTION, TASK_ID) **VALUES** ('Business objectives and success criteria', 'The Business Objectives section describes the goals of the project in business terms. For each objective, Business Success Criteria, i.e., explicit measures for determining whether or not the project succeeded in its objectives, should be provided. This section should also list objectives that were considered but rejected. The rationale of the selection of objectives should be given. ', 1);
INSERT INTO Report_Section(SECTION_NAME, SECTION_DESCRIPTION, TASK_ID) **VALUES** ('Inventory of resources', 'The Inventory of Resources section aims to identify personnel, data sources, technical facilities, and other resources that may be useful in carrying out the project', 1);
INSERT INTO Report_Section(SECTION_NAME, SECTION_DESCRIPTION, TASK_ID) **VALUES** ('Requirements, assumptions, and constraints', 'This section lists general requirements for the project's execution: type of project results, assumptions made about the nature of the problem and the data being used, and constraints imposed on the project.', 1);
INSERT INTO Report_Section(SECTION_NAME, SECTION_DESCRIPTION, TASK_ID) **VALUES** ('Risks and contingencies', 'This section identifies problems that may occur in the project, describes the consequences, and states what actions can be taken to minimize such risks. ', 1);
INSERT INTO Report_Section(SECTION_NAME, SECTION_DESCRIPTION, TASK_ID) **VALUES** ('Terminology', 'The Terminology section allows people unfamiliar with the problems being addressed by the project to become more familiar with them. ', 1);

section_id	section_name	section_descr...	task_id
1	Background	The Background secti…	1
2	Business objectives …	The Business Objecti…	1
3	Inventory of resourc…	The Inventory of Res…	1
4	Requirements, assump…	This section lists g…	1
5	Risks and contingenc…	This section identif…	1
6	Terminology	The Terminology sect…	1
7	Costs and benefits	This section describ…	1
8	Data mining goals an…	The Data Mining Goal…	1
9	Project plan	This section lists t…	1
10	Initial assessment o…	This section gives a…	1

INSERT INTO Technique(TECHNIQUE_NAME, TECHNIQUE_DESCRIPTION, PROBLEM_ID) **VALUES** ('Clustering techniques', 'Clustering techniques consider data tuples as objects. They partition the objects into groups, or clusters, so that objects within a cluster are “similar” to one another and “dissimilar” to objects in other clusters. ', 2);
INSERT INTO Technique(TECHNIQUE_NAME, TECHNIQUE_DESCRIPTION, PROBLEM_ID) **VALUES** ('Neural networks', 'A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.', 2);
INSERT INTO Technique(TECHNIQUE_NAME, TECHNIQUE_DESCRIPTION, PROBLEM_ID) **VALUES** ('Visualization', 'Visualization is any technique for creating images, diagrams, or animations to communicate a message.', 2);

technique_id	technique_name	technique_description	problem_id
1	Clustering techniques	Clustering techniques consider …	2
2	Neural networks	A neural network is a series of…	2
3	Visualization	Visualization is any technique …	2

INSERT INTO Dataset(DATASET_ATTRIBUTE) VALUES ('With missing values');
INSERT INTO Dataset(DATASET_ATTRIBUTE) VALUES ('No missing values');
INSERT INTO Dataset(DATASET_ATTRIBUTE) VALUES ('With noisy values');
INSERT INTO Dataset(dataset_attribute) VALUES ('No noisy values');
INSERT INTO Dataset(DATASET_ATTRIBUTE) VALUES ('Labeled');
INSERT INTO Dataset(DATASET_ATTRIBUTE) VALUES ('Unlabeled');

dateset_id	dataset_attribute	dataset_description
1	With missing values	<null>
2	No missing values	<null>
3	With noisy values	<null>
4	No noisy values	<null>
5	Labeled	<null>
6	Unlabeled	<null>

INSERT INTO Algorithm(algorithm_name, algorithm_description, dataset_id) **VALUES** ('Hierarchical clustering', 'Hierarchical clustering, is based on the core idea of objects being more related to nearby objects than to objects farther away. These algorithms connect "objects" to form "clusters" based on their distance.', 6);
INSERT INTO Algorithm(algorithm_name, algorithm_description, dataset_id) **VALUES** ('k-means clustering', 'In centroid-based clustering, clusters are represented by a central vector, which may not necessarily be a member of the data set. When the number of clusters is fixed to k, k-means clustering gives a formal definition as an optimization problem: find the k cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the cluster are minimized.', 6);
INSERT INTO Algorithm(algorithm_name, algorithm_description, dataset_id) **VALUES** ('Distribution-based clustering', 'The clustering model most closely related to statistics is based on distribution models. Clusters can then easily be defined as objects belonging most likely to the same distribution. ', 6);
INSERT INTO Algorithm(algorithm_name, algorithm_description, dataset_id) **VALUES** ('Density-based clustering', 'In density-based clustering, clusters are defined as areas of higher density than the remainder of the data set. Objects in sparse areas - that are required to separate clusters - are usually considered to be noise and border points.', 6);

algorithm_id	algorithm_name	algorithm_description	dataset_id
1	Hierarchical clustering	Hierarchical clustering, is based …	6
2	k-means clustering	In centroid-based clustering, clus…	6
3	Distribution-based clustering	The clustering model most closely …	6
4	Density-based clustering	In density-based clustering, clust…	6

INSERT INTO Realize(TECHNIQUE_ID, ALGORITHM_ID) VALUES (1, 1);
INSERT INTO Realize(TECHNIQUE_ID, ALGORITHM_ID) VALUES (1, 2);
INSERT INTO Realize(TECHNIQUE_ID, ALGORITHM_ID) VALUES (1, 3);
INSERT INTO Realize(TECHNIQUE_ID, ALGORITHM_ID) VALUES (1, 4);

technique_id	algorithm_id
1	1
1	2
1	3
1	4

INSERT INTO Function(FUNCTION_NAME) VALUES ('linkage');
INSERT INTO Function(FUNCTION_NAME) VALUES ('Euclidean distance');
INSERT INTO Function(FUNCTION_NAME) VALUES ('kMeans');

function_id	function_name	function_description
1	linkage	<null>
2	Euclidean distance	<null>
3	kMeans	<null>

INSERT INTO Employ(ALGORITHM_ID, FUNCTION_ID) VALUES (1, 1);
INSERT INTO Employ(ALGORITHM_ID, FUNCTION_ID) VALUES (2, 2);
INSERT INTO Employ(ALGORITHM_ID, FUNCTION_ID) VALUES (2, 3);

algorithm_id	function_id
1	1
2	2
2	3

Удаления

DELETE FROM table_name **WHERE** [condition];
DELETE FROM Activity **WHERE** task_id=2;

Вывод

Выполнив эту лабораторную работу, я научилась работать с DDL и DML и вспомнила целостности данных.