

MINISTERUL EDUCAȚIEI, CULTURII ȘI CERCETĂRII
Universitatea Tehnică a Moldovei
Facultatea Calculatoare Informatică și Microelectronică
Departamentul Informatică și Ingineria Sistemelor

Admis la susținere

Șef de departament: dr. conf.univ.
Sudacevschi Viorica, conf.univ.dr

„_____” _____ 2021

Dezvoltarea sistemului automatizat de analiză statistică a datelor

Teză de master

Student: _____ (Butacov Dumitru)

Conducător: _____ (Galina Marusic)

Chișinău, 2021

ADNOTARE

La teza de master cu tema “Dezvoltarea sistemului automatizat de analiză statistică a datelor” a
st. gr. MAI-191M Butacov Dumitru

Analiza datelor este utilizată în afaceri pentru ajutarea organizațiilor în lerea deciziilor de afaceri mai bune. Indiferent dacă este vorba de cercetare de piață, cercetare de produs, poziționare, recenzii ale clienților, analiză de sentiment sau orice altă problemă pentru care există date, analiza datelor va oferi informații de care organizațiile au nevoie pentru a lua decizii corecte. Analiza datelor este importantă pentru organizațiile de astăzi, deoarece alegerile bazate pe date sunt singura modalitate de a fi cu adevărat încrezut în deciziile de afaceri. Unele companii de succes pot fi create pe o presimțire, dar aproape toate opțiunile de afaceri de succes sunt bazate pe date. Teza de master este focusată pe studiul și analiza instrumentelor de colectare și analiză a datelor pentru dezvoltarea propriului sistem de automatizat de analiză a datelor statistice.

Teza de master constă din introducere, 3 capitole 15 surse bibliografice, 2 anexe, 28 figuri. În Capitolul 1 „Analiza domeniului de studiu. Necesitatea abordării. Scop și obiective” este efectuat studiul teoretic al domeniului abordat, determinarea actualității și importanța domeniului pentru organizațiile contemporane precum și metode de analiză a datelor de asemenea este defenit scopul și obiectivele acestei teze de master. În Capitolul 2 „Metode statistice privind analiza și vizualizarea datelor” sunt descrise tehnicile generele de analiză a datelor, planul de efectuare a unei analize de date, importanța interpretării datelor corecte și modalitățile de interpretare a acestora precum și metode de vizualizare a datelor și recomandări de utilizare. În Capitolul 3 „Rezultatele cercetării. Studiu de caz” este realizat elaborarea sistemului automatizat, descris procesul de setare a instrumentului de colectare a datelor, stadartizarea, filtrarea agregarea și interpretarea datelor și vizualizarea acestora. Sistemul fiind testat pe studiu de caz și anume pe „Platforma de atașarea a publicității la rezultatele verificării contului de telefoniei mobile prin intermediul comenzii USSD”. Pe baza rezultatelor obținute în studiu de caz a fost generat un raport analitic prin intermediul aplicației Tableau. Concluziile prezintă succint principalele rezultate obținute în lurare.

Cuvinte cheie: *reporting, colectarea datelor, procesarea datelor, analiza datelor.*

ANNOTATION

At the master thesis on topic “ Development of the automated analysis system of statistical data”
of st. gr. MAI-191M Butacov Dumitru

Data analysis is used in business to help organizations make better business decisions. Whether it's market research, product research, positioning, customer reviews, sentiment analysis, or any other data issue, data analysis will provide the information organizations need to make the right decisions. Data analysis is important for today's organizations because data-driven choices are the only way to be truly confident in business decisions. Some successful companies can be created on a premonition, but almost all successful business options are based on data. The master's thesis is focused on the study and analysis of data collection and analysis tools for the development of its own automated system for statistical data analysis.

The master's thesis consists of an introduction, 3 chapters 15 bibliographic sources, 2 appendix, 28 figures. In Chapter 1 “Analysis of the field of study. The need for an approach. Purpose and objectives ”is carried out the theoretical study of the approached field, determining the topicality and importance of the field for contemporary organizations as well as methods of data analysis is also defined the purpose and objectives of this master's thesis. Chapter 2 "Statistical methods for data analysis and visualization" describes the general techniques of data analysis, the plan for performing a data analysis, the importance of correct data interpretation and how to interpret them, as well as methods of data visualization and recommendations. use. In Chapter 3 “Research Results. Case study ”is performed the development of the automated system, described the process of setting the data collection tool, staging, filtering aggregation and interpretation of data and viewing them. The system is being tested on a case study, namely on the "Platform for attaching advertising to the results of checking the mobile phone account through the USSD command". Based on the results obtained in the case study, an analytical report was generated through the Table application. The conclusions briefly present the main results obtained in the study.

Keywords: *reporting, data extract, data processing, data analysis.*

CUPRINS

INTRODUCERE	8
1 ANALIZA DOMENIULUI DE STUDIU. NECESITATEA ABORDĂRII. SCOP ȘI OBJECTIVE	10
1.1. Definirea conceptului de analiza datelor și importanța acestuia	10
1.2. Metode de analiză a datelor	12
1.3. Conceptul de „big data” și formarea acestuia	14
1.4. Metode de analiză a datelor	16
1.5. Scopul și obiectivele tezei	20
2 METODE STATISTICE PRIVIND ANALIZA ȘI VIZUALIZAREA DATELOR	21
2.1. Tehnici de analiză a datelor	21
2.2. Interpretarea datelor	26
2.3. Vizualizarea datelor și informațiilor	31
3 REZULTATELE CERCETĂRII. STUDIU DE CAZ	39
3.1. Studiul și analiza instrumentelor de analiză statistică a datelor	39
3.2. Colectarea datelor	40
3.3. Procesarea datelor	45
3.4. Vizualizarea datelor	48
CONCLUZII	51
BIBLIOGRAFIE	52
ANEXĂ 1 COD SURSĂ	53
ANEXĂ 2 CALCULĂRI TABLEAU	58

INTRODUCERE

Analiza datelor nu este doar prelucrarea informațiilor după ce a fost primită și colectată, ci este un mijloc de testare a ipotezelor. Scopul oricărei analize a datelor este de a înțelege întreaga situație studiată (identificarea tendințelor, inclusiv abaterile negative de la plan, prognozarea și primirea recomandărilor). Pentru a atinge acest obiectiv, sunt stabilite următoarele sarcini de analiză a datelor[2]:

- colectarea de informații;
- structurarea informațiilor;
- identificarea tiparelor, analize;
- prognozarea și primirea recomandărilor.

Toate datele conțin informații importante, dar pentru probleme diferite. Pentru a selecta date utile pentru o anumită situație din întregul masiv de date, este necesară prelucrarea lor. Cu o cantitate mare de date, o astfel de analiză nu ar trebui să aibă loc în Excel, ci într-un program analitic profesional.

Soluțiile de analiză variază în funcție de costul implementării și de personalul care va fi utilizat pentru a le gestiona. Există soluții pentru corporațiile mari în care nu se poate face fără investiții mari - atât în etapa de implementare, cât și ulterior în etapa de lucru a soluției (cheltuielile principale sunt specialiștii care lucrează cu software). Pentru companiile mijlocii și mici, există soluții precum Tableau: costul implementării acestora este minim (de 840 de USD anual - la începutul anului 2021), nu necesită cunoștințe speciale - orice utilizator cu cunoștințe inițiale de calculator poate face un frumos reportaj interactiv.

O echipă ce se ocupa cu analiza datelor permite la dezvoltarea și implementarea soluțiilor pentru diverse industrii și domenii de afaceri:

- analiza vânzărilor;
- prognozarea eficacității promoțiilor;
- urmărirea canalului de vânzări pentru o anumită campanie cu rate de conversie în fiecare etapă;
- segmentarea prin diferite metode: ABC, RFM etc. ;
- indicatori cheie ai marketingului pe internet;
- analiza mărfurilor în depozite;
- analize pentru departamentul de resurse umane (carduri de angajați, instruire, analiză KPI etc.) ;
- analiza indicatorilor financiari;
- și mult mai mult.

Reieșind din cele menționate a fost formulat scopul tezei:

Studiul și analiza modelelor și metodelor ce vor sta la baza unui sistem informațional ce urmărește scopul de colectare, analiză, filtrare, prelucrare și vizualizare a datelor.

Pentru atingerea scopului, au fost formulate următoarele **obiective**:

1. Studiul și analiza instrumentelor, metodelor și modelelor de colectare, analiză, prelucrare a datelor;
2. Elaborarea tehnicilor de prelucrare și interpretare a datelor;
3. Utilizarea aplicației SymmetriDS în scopul coletării datelor;
4. Utilizarea bazei de date relaționale MySQL pentru păstrarea datelor;
5. Aplicarea în practică a motorului de interogare SQL Presto și a limbajului de programare Python pentru analiza, filtrarea și agregarea datelor;
6. Interpretarea, analizarea și vizualizarea datelor prin intermediu aplicației Tableau.

BIBLIOGRAFIE

1. Medotica analizelor de date. Disponibil: <https://loginom.ru/blog/base-ml>
2. Analist de date: abilități, responsabilități. Disponibil: <https://checkroi.ru/blog/professiya-analitik-dannyh/>
3. Ce este analiza datelor. Disponibil: <https://www.investopedia.com/terms/d/data-analytics.asp>
4. Soluții realizate ale analizei de date. Disponibil: <https://asu-analitika.ru/>
5. Știința datelor: termeni, aplicație, educație. Disponibil: <https://netology.ru/blog/01-2020-gid-po-data-science>
6. Analiza și gestionarea datelor pentru decizii. Disponibil: <https://www.cio.com/article/3606151/what-is-data-analytics-analyzing-and-managing-data-for-decisions.html>
7. Cele mai utile metode de analiză a datelor. Disponibil: <https://careerfoundry.com/en/blog/data-analytics/data-analysis-techniques/>
8. Ce este analiza datelor și de ce contează. Disponibil: <http://helpex.ru/programmirovanie/chto-takoe-analiz-dannyh-i-pochemu-jeto-vazhno>
9. Ce face un analist de date. Disponibil: <https://www.northeastern.edu/graduate/blog/what-does-a-data-analyst-do/>
10. Cum a început era Big Data. Disponibil: <https://rb.ru/story/era-big-data/>
11. Ce este analiza datelor. Disponibil: <https://www.investopedia.com/terms/d/data-analytics.asp>
12. Analiza datelor în bussinesul modern. Disponibil: <https://www.datapine.com/blog/data-analysis-methods-and-techniques/#:~:text=The%20descriptive%20analysis%20method%20is,valuable%20insights%20to%20your%20business.>
13. Ce este analiza datelor. Disponibil: <https://hackr.io/blog/what-is-data-analysis-methods-techniques-tools>
14. Analiza datelor în cercetare. Disponibil: <https://www.questionpro.com/blog/data-analysis-in-research/>
15. Ce este vizualizarea datelor. Disponibil: <https://www.klipfolio.com/resources/articles/what-is-data-visualization>
16. Beneficiile și problemele interpretării datelor. Disponibil: <https://www.datapine.com/blog/data-interpretation-methods-benefits-problems/>