

TALLINNA TEHNIKAÜLIKOOL  
Infotehnoloogia teaduskond

Johanna Brigitte Verro

**Andmete analüüsi ja valimi kontrollimise  
automatiseerimine ettevõtte Äripäeva AS  
konkurentsiraportite näitel**

Bakalaureusetöö

Juhendaja: Inna Švartsman, MSc

Tallinn 2023

## **Autorideklaratsioon**

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Johanna Brigitte Verro

17.05.2023

## **Annotatsioon**

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 30 leheküljel, 4 peatükki, 10 joonist, 1 tabelit.

## **Abstract**

### **Automating Data Analysis and Sample Validation Using the Example of Äripäev AS Reports**

The thesis is in Estonian and contains 30 pages of text, 4 chapters, 10 figures, 1 table.

## Lühendite ja mõistete sõnastik

MTA	Maksu- ja tolliamet
EMTAK	Eesti Majanduse Tegevusalade Klassifikaator, rahvusvaheliselt ühtlustatud Euroopa Ühenduse majandustegevusalade statistilise klassifikaatori (NACE) Eesti rahvuslik versioon
VBA	<i>Visual Basic Application</i> , programmeerimiskeel

## Sisukord

Autorideklaratsioon.....	2
Annotatsioon.....	3
Abstract Automating Data Analysis and Sample Validation Using the Example of Äripäev AS Reports .....	4
Lühendite ja mõistete sõnastik .....	5
Sisukord.....	6
Jooniste loetelu .....	8
Sissejuhatus .....	9
Taust.....	9
Probleem.....	9
Eesmärk .....	10
Töö struktuur .....	10
1 Varasem ülevaade .....	11
1.1 Konkurentsi raportite olemus.....	11
1.2 Konkurentsi raportite koostamine .....	11
1.3 Valimi kontrollimise ja andmeanalüüsi automatiseerimine .....	13
2 Metoodika.....	16
2.1 Ülevaade objektist.....	16
2.1.1 Esialgset nõuded .....	16
2.1.2 Ettevõtte valimisse sobivuse nõuded.....	17
2.2 Tööriistadest mida kasutatakse .....	18
2.3 Protsessist kuidas tehti.....	19
3 Peamised tulemused.....	21
3.1 Nõuded.....	21
3.2 Arhitektuur ja disain.....	24
3.2.1 Esialgse valimi koostamine.....	25
3.2.2 Manuaalne kontroll.....	26
3.2.3 Lõpliku valimi koostamine .....	27
3.2.4 Andmete kogumine.....	27

3.2.5 Andmetabeli koostamine .....	29
3.3 Kasutajaliides .....	30
4 Analüüs ja järeldused .....	35
4.1 Tehnoloogia valik.....	35
4.2 Efektiivsus ja täpsus .....	35
4.3 Põhjendus.....	37
4.4 Kitsaskohad ja edasiarendused .....	38
5 Kokkuvõte .....	40
Kasutatud kirjandus .....	41
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks .....	43

## Jooniste loetelu

Joonis 1. Otsustuspuu ettevõtte valimisse sobivuse hindamisel.....	18
Joonis 2. Ülevaade rakenduses kasutatavatest moodulitest. ....	24
Joonis 3. Suhtlusdiagramm esimese valimi koostamise jaoks. ....	26
Joonis 4. Jadadiagramm andmete kogumise sammude näitamiseks.....	28
Joonis 5. Suhtlusdiagramm andmetabeli koostamise makrost. ....	29
Joonis 6. Kasutajaliides lehel “Input”.....	31
Joonis 7. Kasutajaliidese vaade lehel “InitialSample”.....	31
Joonis 8. Kasutajaliides lehel “ManualCheckup”.....	32
Joonis 9. Kasutajaliides lehel “Sample”.....	33
Joonis 10. Kasutajaliidese vaade lehel “Data”. ....	33
Joonis 11. Ülevaade efektiivsus testide tulemustest. ....	36



## **Sissejuhatus**

### **Taust**

Äripäeva infopank on Eesti üks juhtivaid ettevõtete andmebaase, mis sisaldab laia valikut majandusandmeid ning ettevõtete ja organisatsioonide profiile. Infopangas on kättesaadavad andmed ettevõtete finantsseisundi, maksuvõla, omanike, juhatuse liikmete ja muu olulise kohta. Äripäeva infopank pakub ühe tootena konkurentsiraporteid, mis annavad ülevaate konkreetse sektori ettevõtete finantsnäitajatest, turuosast ja -positsioonist. Konkurentsiraportid on seega väärtuslikud tööriistad, mis aitavad ettevõtetel teha strateegilisi otsuseid ning jälgida nende tegevusvaldkonna turuolukorda.

### **Probleem**

Ettevõtte kogub andmeid erinevatest allikatest ja tihti on andmed salvestatud erinevates formaatides ja andmebaasides, mis muudab andmete kättesaadavuse ja töötlemise keeruliseks. Konkurentsiraportite koostamine toimub Excelis käsitsi ning see nõuab palju käsitsi tehtavaid andmebaasi päringuid ja andmete kopeerimist. Käsitsi päringutele sisendite andmine, Excelisse valemite käsitsi sisestamine ja andmete kopeerimine suurendab inimlike vigade tõenäosust ja on tülikas ning aeganõudev protsess. Konkurentide ees eelise saamiseks on vaja saada raportid välja võimalikult kiiresti ja selletõttu on ühe raporti koostamisel ajaline piirang. See omakorda vähendab aega valimi kontrollimiseks ja võib viia sama valimi mitmekordse kasutamiseni, ilma et seda oleks ajakohastatud. Praeguses keerulises ja pidevalt muutuv majanduskeskkonnas võib ettevõtete tegevusvaldkond ja suund muutuda kiiremini, kui on võimalik korrigeerida valimit. Seetõttu on vajalik täpsustada valimi koostamist ja andmete analüüsi, et tagada kvaliteetsem ülevaade sektori turuolukorrast, kasutades samal ajal kõige ajakohasemaid andmeid.

## **Eesmärk**

Lõputöö eesmärgiks on leida võimalusi, kuidas automatiseerida valimi koostamist ja andmete analüüsi, et suurendada Äripäeva infopanga konkurentsiraportite kvaliteeti ja aidata klientidel paremini langetada strateegilisi otsuseid. Selleks on vaja välja töötada protsessi automatiseerimise lahendus, mis aitaks suurendada täpsust, samal ajal vähendades ettevõtte aja- ja ressursikulu. Lõpptulemusena peaksid konkurentsiraportid olema täpsemad ja usaldusväärsemad, mis võimaldab klientidel teha paremaid strateegilisi otsuseid.

## **Töö struktuur**

Autor annab ülevaate projekti praegust seisundit, identifitseerib vajalikud muudatused eesmärkide täitmiseks ning analüüsib varasema kirjanduse ülevaate ning ettevõtte nõuete põhjal erinevaid võimalusi nende eesmärkide saavutamiseks. Seejärel leiab autor parima lahenduse probleemidele ning koostab prototüübi. Pärast töö valmimist hindab võimalusi projekti edasiseks arendamiseks.

# **1 Varasem ülevaade**

## **1.1 Konkurentsi raportite olemus**

Konkurentsiraport võimaldab võrrelda ettevõtteid erinevate majandusnäitajate alusel, andes ülevaate nende finantsnäitajatest, maksuvõlast, juhatuse liikmetest ja muudest olulistest faktoritest. Raportid võimaldavad hinnata ettevõtete finantsseisu ja paremusjärjestust erinevate finantsnäitajate alusel. Analüüsitakse peamiselt müügitulu, makstud tööjõumakse, töötajate arvu ja erinevaid nendest välja arvatatud suhtarve. Olulised on näiteks käive töötaja kohta, brutopalga paremusjärjestus ja trendimuutused. Edetabelid koostatakse nii eelmise kvartali kui ka nelja kvartali keskmise põhjal kõigi näitajate kohta.

Konkurentsiraporti koostamisel on oluline tagada võimalikult värskete andmete kasutamine, seega eelistatakse viimaseid deklareeritud andmeid. Raportit tehes kasutatakse viimase kvartali andmeid ning esitamise tähtaja järgi viimast majandusaasta aruannet. Tuleb arvestada, et majandusaasta aruannete esitamise tähtaeg (30. juuni) [1] mõjutab raporti koostamist ning kõik enne 30. juunit valmivad raportid kasutavad 2021.a majandusaasta aruannet. Peale seda valmivad raportid aga juba kasutades 2022.a majandusaasta aruannet.

Turuseisu korrektseks hindamiseks on oluline määratleda raporti aluseks olev valim. Kõik Eestis tegutsevad ettevõtted on registreeritud kindla tegevusala järgi, mida tähistatakse Eesti Majandustegevuse Klassifikaatoriga (EMTAK) [2]. EMTAK liigitab ettevõtted vastavalt nende tegevusaladele rühmadesse ja alarühmadesse ning on oluline allikas valdkonnastatistika jaoks [3]. Seetõttu kasutatakse ka konkurentsiraportites EMTAK-i valimi koostamisel.

## **1.2 Konkurentsi raportite koostamine**

Konkurentsiraportite koostamiseks kasutatakse Exceli 2013.a versiooni ning igale kasutajale meelepärast andmebaasi rakendust. Kõige populaarsem rakendus andmebaasi

päringute tegemiseks on SQL Server Management Studio (SSMS). Tegu on väga võimeka tarkvara, mis võimaldab hallata ja konfigureerida SQL Serverit. SSMS-i saab kasutada andmebaasi päringute tegemiseks, skriptide genereerimiseks, andmete importimiseks ja eksportimiseks ning tulemuste visualiseerimiseks [4].

Konkurentsiraportite koostamine nõuab hoolikat ettevalmistust ja põhjalikku tööd. Esimeseks sammuks on teha andmebaasipäring, et saada esialgne valim. Seejärel võrreldakse esialgse valimi tulemusi eelmise valimiga, et selgitada välja, millised ettevõtted vajavad käsitsi ülevaatamist. Kõik uued ettevõtted, keda eelmises valimis ei olnud, kontrollitakse käsitsi üle, et selgitada, kas need sobivad valimisse või mitte. Seejärel koostatakse lõplik valim, mida kasutatakse raporti koostamiseks. Pärast seda tehakse andmebaasipäringud, et saada raporti jaoks puudu olevad andmed. Alguses salvestatakse kõik andmed erinevatele lehekülgedele, kuid hiljem pannakse need kokku ühte tabelisse, kasutades erinevaid Exceli valemeid.

Selle meetodika kasutamisel on mitmeid plusse ja miinuseid. Üks suurimaid eeliseid on võimalus tulemusi igas etapis kontrollida. Kuna kõik toimingud tehakse käsitsi, siis on kasutajal hea ülevaade sellest, mis parasjagu toimub ning kui tekib olukord, kus andmeid vaadates tundub, et kuskil võib viga olla, saab seda kiiresti kontrollida. See võib olla eriti oluline, kui ettevõtted on deklareerinud oma igakuise töötajate arvu või müügitulu valesti. Näiteks, kui ettevõtte käive on olnud iga kvartal alla 50 000 euro ning järsku üks kvartal näitab tulu 1,5 miljonit eurot, siis on suur tõenäosus, et midagi on deklareeritud valesti ning seda ei saa raportisse lisada enne, kui viga on kõrvaldatud.

Käsitsi läbiviidava töö suurim puudus on andmete ebakorrektsus, mis võib tuleneda inimlikest vigadest. Andmete käsitsi sisestamine ja töötlemine võib olla aeganõudev ja tülikas protsess, mis suurendab vigade tekkimise riski. Näiteks võib valesti sisestatud valem või valesti kopeeritud andmed viia ebaõigete tulemusteni. Inimlikud eksimused võivad ilmneda juba päringute koostamisel ja nende sisestamisel, mis võib põhjustada vigu juba esialgses valimis. Kui viga ei tuvastata õigeaegselt on raport, mis põhineb vigasel valimil, kasutu.

Selle meetodika suurimaks puuduseks on ajakulu. Ühe valdkonnaga raporti loomine võtab aega umbes 3 tundi, sellest 2 tundi läheb andmepäringute tegemiseks ja tulemuste vormistamiseks ühisesse tabelisse. Umbes tund aega kulub käsitsi valimi kontrollimiseks.

See tähendab, et mitme valdkonnaga raportite loomiseks kulub veel vähemalt tund aega lisaks, sest mitme valdkonna puhul on käsitsi kontrolli vajavate ettevõtete hulk kordades suurem. See on aga aeg, mida ei ole kuskilt võtta. Selle tulemusena võib suurte valimite puhul juhtuda, et puudub aeg valimit põhjalikult kontrollida ning võetakse kasutusele eelmise kvartali valim.

### **1.3 Valimi kontrollimise ja andmeanalüüsi automatiseerimine**

Automatiseerimine parandab aruannete loomise ja levitamise tõhusust ja efektiivsust. EY raporti järgi on võimalik ligikaudu 65% inimressurssidega seotud ülesannetest, sealhulgas andmete puhastamisest, automatiseerida [5]. Mida suurem osa andmetöötlusest on automatiseeritud, seda väiksem tõenäosus on teha inimvigu.

Andmetöötlus Excelis on läbipaistev, mis tähendab, et kui Exceli faili avada, siis on kohe nähtavad andmed ja igasugune andmetöötluse samm. See annab ülevaate, kuidas andmetega manipuleeritakse ja kuidas tulemusi saavutatakse. Teised tarkvaratööriistad ja programmeerimiskeeled aga varjavad paljusid andmetöötluse aspekte, mis võib olla kasulik edasijõudnud kasutajatele, kuid algajatele ja visuaalsetele õppijatele ei ole see kasulik. Seetõttu on Excel kasulik tööriist algajatele, kes soovivad saada põhjalikku ülevaadet andmetöötlusest [6].

Järgnevalt annab autor ülevaate mõningatest programmeerimiskeeltest, mida saab kasutada Excelil põhineva andmeanalüüsi automatiseerimiseks.

VBA ehk Visual Basic for Applications on programmeerimiskeel, mida kasutatakse laialdaselt Microsoft Office'i rakenduste, eriti Exceli, automatiseerimiseks ja kohandamiseks. VBA võimaldab kasutajatel luua ja käivitada skripte, mis saavad automatiseerida korduvaid ülesandeid ja töödelda suuri andmekogumeid. Üks võimalus selleks on luua makrosid. Makrod on VBA "programm", mida saab käivitada Excelis [7]. Lisaks on VBA-s võimalik luua "töölehe-funktsioon", mida saab kasutada nagu Excelisse sissekirjutatud funktsioone näiteks "SUM" või "MEAN" [7] [8].

VBA on laialdaselt kasutusel ettevõtetes ja organisatsioonides, eriti finantsvaldkonnas, kus Excel on laialdaselt kasutatav andmete töötlemise ja analüüsi tööriist [7]. VBA eelis on veel, et tegu on kompileeritud keelega, mis muudab selle võrreldes skriptikeelega iseseisvates rakendustes kiiremaks [9].

R on programmeerimiskeel, mis on spetsiaalselt loodud statistiliste arvutuste ja graafika loomise jaoks. See on laialdaselt kasutusel andmeanalüüsis, andmeteaduses, statistikas ja teadusuuringutes. R Studio on võimas tööriist, mis võimaldab kasutajatel R-i programmeerimiskeelt kasutada ning tulemusi kuvada ja analüüsida [10]. R Studio saab hakkama nii struktureeritud, kui struktureerimata andmetega [11].

R ja Exceli vaheline seos tugineb kahe osa ühendamisele: RExcel on Microsoft Exceli lisandmoodul, mis pakub R-i ja statconnDCOM-i tabelilaadset liidest. statconnDCOM võimaldab iga Windowsi rakendusel kasutada R-i funktsionaalsust läbi lihtsalt kasutatava COM-komponendi. RExcel võimaldab luua VBA-le sarnaselt uusi funktsioone ja erinevaid makrosid [8].

Python on suhteliselt lihtne üldotstarbeline programmeerimiskeel, mida saab kasutada mitmesuguste ülesannete jaoks, sealhulgas andmeanalüüsiks [12] [13]. Tegu on väga populaarse ja endiselt arenduses oleva keelega [9]. Pythoni on mitmeid analüüsi jaoks sobilikke pakette nagu näiteks NumPy, Pandas ja SciPy. Pythonil on ka palju kasutajasõbralikke andmestruktuure ja võimekaid graafikafunktsioone. Paketi Pandas abil on võimalik andmeid lugeda mitmest erinevast allikast ning salvestada neid erinevatesse faili formaatidesse [12].

Excelisse on võimalik integreerida ka Pythoni. Seda saavutatakse, võimaldades kasutajatel kutsuda Pythoni funktsioone otse Exceli tabelarvutusprogrammis. Üheks näiteks on pakett PyXLL [pyx], mis võimaldab luua Exceli lisandmoduleid Pythonis, mitte VBA-s. Pythoni funktsioonid saab seega teha kättesaadavaks kui töölehtede funktsioonid või makrod. Veel üks sarnane pakett on xlwings [xlw], mis võimaldab samuti automatiseerida Excelit kasutamata selleks VBA-d [13].

Pythoni rakendamine on mõnel juhul sama lihtne kui VBA puhul, näiteks juurdepääs XML-andmetele. Teised elemendid, nagu paketid numbriliste arvutuste jaoks, muudavad Pythoni rakendamise efektiivsemaks. Eelpool mainitud tugevused annavad Pythonile kahtlemata eelise standardse VBA ees [9].

Nii Pythoni kui R programmeerimiskeelega on võimalik teha veel tekstikaevandamist. Tekstikaevandus ehk tekstiandmete kaevandamine või teadmiste avastamine tekstipõhistest andmebaasidest viitab üldiselt teadmiste väljavõtmisele struktureerimata tekstidokumentidest [14].

Pythoni ja R programmeerimiskeelte andmeanalüüsi sammud erinevad üksteisest ning visuaalselt esitatud andmete selgus on Pythoni keeles parem. Python on R programmeerimiskeelega võrreldes kiirem. Pythoni ja R programmeerimiskeelte erinevus seisneb nende lähenemisviisides andmete süstematiseerimisel uurimistegevustes. Python pakub kasutajasõbralikumad graafilist kasutajaliidest, kuid R programmeerimiskeel ei oma neid omadusi, mis mõjutab mõlema platvormi arvutuslikke esitusvõimeid [15].

## 2 Metoodika

### 2.1 Ülevaade objektist

Äripäeva Infopangas pakutakse mitmeid erinevaid tooteid, sealhulgas ka konkurentsi raporteid. Nende raportite peamine eesmärk on anda ülevaade turul toimuvast ning neid koostatakse erinevates valdkondades, et anda sektorist üldine pilt. Raportid sisaldavad olulisi finantsandmeid ettevõtete kohta. Nende andmete põhjal koostatakse erinevaid graafikuid ning tehakse paremusjärjestusi käibe, palga ja töötajate arvu alusel. Kiire raportite valmimine ning nende kvaliteet on aga äärmiselt oluline ning seetõttu on vajalik protsess automatiseerida. Autor loob tarkvaralahenduse, mis automatiseerib andmete pärimise andmebaasist, nende analüüsimise ja vastavusse viimise ning valimi koostamise. See võimaldab oluliselt kiirendada algandmete faili valmimist ning muudab raporti koostamise protsessi tõhusamaks ja efektiivsemaks. Tarkvaralahendus vähendab ka inimlike vigade riski, mis võib esineda käsitsi andmete töötlemise korral, tagades seeläbi raporti kõrgema kvaliteedi. Kokkuvõttes aitab autor loodav tarkvaralahendus kaasa konkurentsi raportite koostamise protsessi täiustamisele.

#### 2.1.1 Esialgsed nõuded

Tehnilised nõuded on spetsiifilised nõuded, mis kirjeldavad tarkvara, süsteemi või seadme tehnilisi omadusi ja funktsioone. Need nõuded hõlmavad tavaliselt riistvara- ja tarkvarakomponente, võrgu- ja andmebaasiühendusi, andmete kaitsemeetmeid ja muid tehnilisi aspekte, mis on vajalikud tarkvara, süsteemi või seadme nõuetekohaseks toimimiseks. Tehnilised nõuded on olulised, et tagada süsteemi turvalisus, stabiilsus ja jõudlus ning et tagada süsteemi ühilduvus teiste süsteemide ja rakendustega, mida võidakse kasutada samas keskkonnas. Tehnilised nõuded peavad olema selged ja täpsed ning vastama kehtestatud standarditele ja parimatele tavadele, et tagada nende täitmise lihtsus ja tõhusus.

Esialgsed tehnilised nõuded tarkvarale:

- Lahendus peab olema jooksvatav kõikidest Äripäeva töötajate arvutitest ehk ei vaja spetsiifilist programmi selle jooksumiseks. See tagab, et kõik saavad programmi kasutada, mitte ainult need, kellel on olemas kindel tarkvara.



- Tarkvara peab suutma töödelda suuri andmekogumeid tõhusalt ja kiiresti, sest mitmete valdkondade valimid on väga suured.
- Tarkvara peab olema skaleeritav, nii et see suudaks toime tulla tulevikus veelgi suuremate andmekogumitega, sest mida rohkem toodet edasi arendatakse, seda rohkemaid sektoreid käsitletakse ja seda suuremaks võivad muutuda valimid.
- Programmi tulemuste aruandlus peab olema täpne, kuvades kõik vajalikud andmed selgelt ja arusaadavalt.

Ettevõtte tarkvara arenduse ärinõuded on määratletud klientide vajaduste rahuldamiseks ja nende töö hõlbustamiseks. Nõuded sisaldavad tarkvara funktsionaalsust, kasutusmugavust, dokumentatsiooni ja andmete esitamist, mis kõik on olulised tarkvara tõhusa kasutamise ja kasutajate rahulolu tagamiseks.

Esiolulised nõuded:

- Rakenduse kasutamine peab võtma vähem aega kui läheks raporti käsitsi koostamisele, sest aeg, mis võib kuluda rapordi tegemisele on piiratud.
- Lõplik valim peab olema nähtav esialgselt valimist eraldi, et valimi ettevõtteid üle vaadates ei peaks neid filtreerima suurest valimist.
- Kasutajaliides peab olema kasutajasõbralik ja intuitiivne, võimaldades lihtsat navigeerimist erinevate funktsioonide vahel.
- Andmed tuleb kuvada selgelt ja arusaadavalt, nii et kasutajad saaksid kiiresti leida otsitava teabe
- Tarkvara peab olema disainitud nii, et kasutajad saaksid andmeid lihtsalt eksportida erinevatesse vormingutesse või analüüsimitarkvarasse.

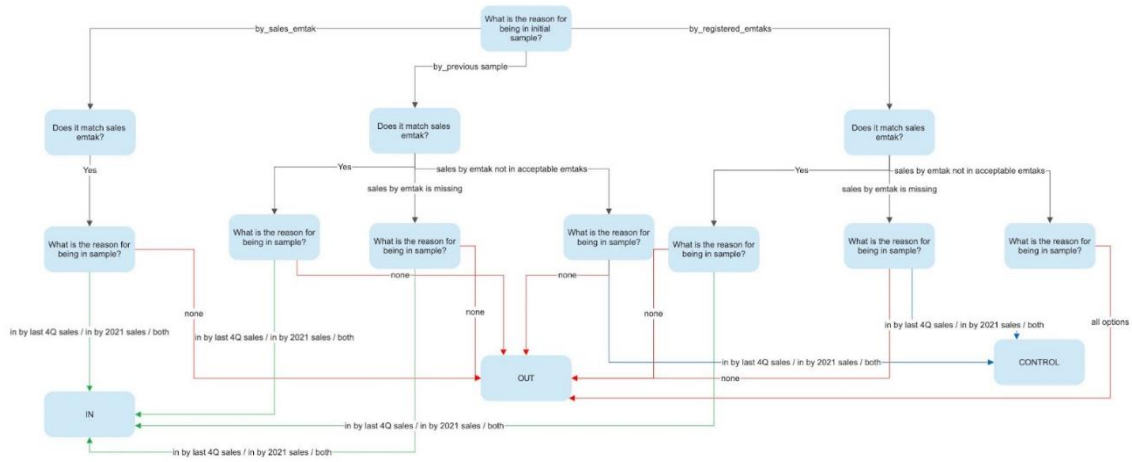
### **2.1.2 Ettevõtte valimisse sobivuse nõuded**

Adekvaatse turu ülevaate saamiseks on kõige alus korrektne valim. Ettevõtte sobivuseks valimisse tuleb täita mitmeid tingimusi lisaks käibepiiri ületamisele. Vastavalt erinevatele tingimustele vaadatakse kas ettevõtte läheb automaatselt lõppvalimisse, jääb välja või läheb manuaalsesse kontrolli.

Esmalt kontrollitakse, kuidas ettevõtte esialgsesse valimisse valiti. Kui ettevõtte valiti majandusaasta müügitulu jaotuse alusel määratud EMTAK-i järgi, on otsustamine lihtne - kuna müügitulu EMTAK on sobilik, tuleb kontrollida vaid, kas käibepiir sai ületatud. Kui jah, siis ettevõtte kuulub automaatselt valimisse; kui mitte, jääb ettevõtte välja.

Ettevõtete jaoks, mis said esialgsesse valimisse eelmise kvartali valimi või ettevõtte poolt registreeritud EMTAK-i kaudu, on otsustamine keerukam. Kui käibepiir ei ole ületatud, siis ettevõtte läheb automaatselt välja kõigis olukordades. Kui ettevõtte pärineb eelmisest valimisest ja käibepiir on ületatud, kuid registreeritud EMTAK ei ole lubatud EMTAK-ide hulgas, siis läheb ettevõtte kontrolli. Kui müügitulu jaotuse EMTAK on puudu, siis usaldame eelmise korra kontrolli ja võtame ettevõtte automaatselt valimisse, arvestades valdkonna raporti lisamise põhjust. Kui ettevõtte kuulub lubatud EMTAK-ide hulka lisaks käibepiir ületamisele, siis ettevõtte läheb automaatselt lõplikku valimisse.

Ettevõtete puhul, kes on valitud ettevõtte poolt registreeritud EMTAK-i järgi, läheb ettevõtte automaatselt valimisse ainult siis, kui EMTAK kuulub lubatud EMTAK-ide hulka ja käibepiir on ületatud. Kui käibepiir ületatakse, kuid EMTAK ei kuulu lubatud EMTAK-ide hulka, siis läheb ettevõtte manuaalsesse kontrolli. Teistel juhtudel läheb automaatselt välja.



Joonis 1. Otsustuspuu ettevõtte valimisse sobivuse hindamisel.

## 2.2 Tööriistadest mida kasutatakse

On palju erinevaid tööriistu, mille abil saab teostada erineva keerukusastmega andmeanalüüsi. Peamise tööriista valikul sai määravaks konkreetse ettevõtte soov, et

kasutajad ei peaks selle jaoks alla laadima eraldi tarkvara. Lähtudes sellest sai valitud konkurentsiraportite andmeanalüüsi automatiseerimiseks VBA. VBA (Visual Basic for Applications) on programmeerimiskeel, mis on spetsiaalselt loodud Microsofti Office'i programmide jaoks. VBA kasutamine Excelis võimaldab luua automatiseeritud töövahendeid ja rakendusi, mis aitavad suurendada töö efektiivsust ja täpsust.

Kui ettevõtte soovib automatiseerida oma analüüsi protsessi Excelis, on VBA tavaliselt parim valik, kuna see võimaldab luua kompleksseid skripte ja makrosid, mis teevad palju tööd kiiresti ja täpselt ilma eraldi tarkvara alla laadimata. Lisaks on VBA tugevasti integreeritud Exceli keskkonda, mis muudab selle kasutamise lihtsaks ja mugavaks.

Lisaks on VBA-ga loodud lahenduste hooldamine ja täiendamine tavaliselt lihtsam kui muude programmeerimiskeeltega. Tänu sellele on Excelis andmete töötlemine ja analüüsimine on lihtsam ning võimaldab saavutada paremaid tulemusi ja suurendada töö efektiivsust

## **2.3 Protsessist kuidas tehti**

Tarkvaralahenduse loomine võib suures pildis jagada neljaks etapiks: tutvumine olemasoleva lahendusega, uurimine ja planeerimine, nõuete paikapanek ning prototüübi loomine ja analüüs.

Esimene etapp on tutvumine olemasoleva lahendusega. Autor alustas sellega, et tutvus konkurentsi raportite koostamise süsteemiga, mida soovitakse parandada või uuendada. Selleks kaardistas ta esialgse seisundi, mis sisaldas ülevaadet sellest, milliseid programme kasutati, milline oli ajakulu ning mis olid olemasoleva programmi suurimad kitsaskohad. Paika sai pandud ka ettevõtte pooled nõuded. See annab autorile vajaliku konteksti uue lahenduse loomiseks ja parendamiseks. Lisaks aitab see tuvastada probleeme, mida uus tarkvaralahendus peaks lahendama.

Teine etapp on uurimine ja planeerimine. Selles etapis uuris autor varasemat kirjandust ja tehnoloogiaid, mis võiksid aidata uue lahenduse loomisel. Selles etapis pani autor ka paika prototüübis kasutatavad vahendid: VBA ja Excel. Autor koostas plaani uue prototüübi loomiseks, mis hõlmas automatiseerimist kolmes osas. Esimeses osas planeeriti automatiseerida andmete analüüsimine ja lõpptulemuste tegemine. Selles osas oli andmebaasipäringuid veel vaja teha käsitsi. Teises osas planeeriti automatiseerida

andmebaasipäringute tegemine. Lõpuks kolmandas osas viidi need kaks osa kokku ning tehti vastavad muudatused süsteemide ühildamiseks.

Kolmas etapp on nõuete paika panemine. Autor vestles erinevate inimestega, et saada teada, millised on nende vajadused ja ootused uuele lahendusele. Selle põhjal sai luua tehnilised ja ärilised nõuded, mis aitasid autori meeskonnal selgitada, millised funktsioonid on uues rakenduses vajalikud ja millised mitte.

Viimane etapp on prototüübi loomine ja analüüs. Autor valmistas prototüübi ette vastavalt plaanile. Kui prototüüp oli valmis, siis analüüsiti selle tulemusi ja võrreldi neid algsete nõuetega. Autor tegeles vajalike muudatuste ja paranduste tegemisega, et jõuda soovitud tulemusteni.

## 3 Peamised tulemused

Valminud lahendus on leitav SIIT.

### 3.1 Nõuded

Peale olemasoleva lahendusega tutvumist ning ettevõttepoolsete nõuete kaardistamist sai autor luua plaani uue tarkvaralahenduse loomiseks. Selleks pani autor paika järgnevad nõuded.

Tehnilised nõuded:

- Analüüsi protsess peab koosnema omavahel sõltumatutest osadest, et võimaldada manuaalsete andmekorrektsioonide tegemist vajadusel. See on eriti oluline juhul, kui mõned majandusaasta aruanded ei ole masinloetavad ning nende andmeid tuleb erinevates analüüsi etappides käsitsi korrigeerida.
- Analüüsi protsessi erinevaid osi tuleb saada käivitada eraldi ning vajadusel ka üksikuid juba varem läbitud samme korrata, ilma et peaks kogu protsessi uuesti alustama. Selleks peab protsess olema hästi struktureeritud ning iga osa peab olema iseseisvalt toimiv, ilma et see mõjutaks teisi osi.

Ärilised nõuded:

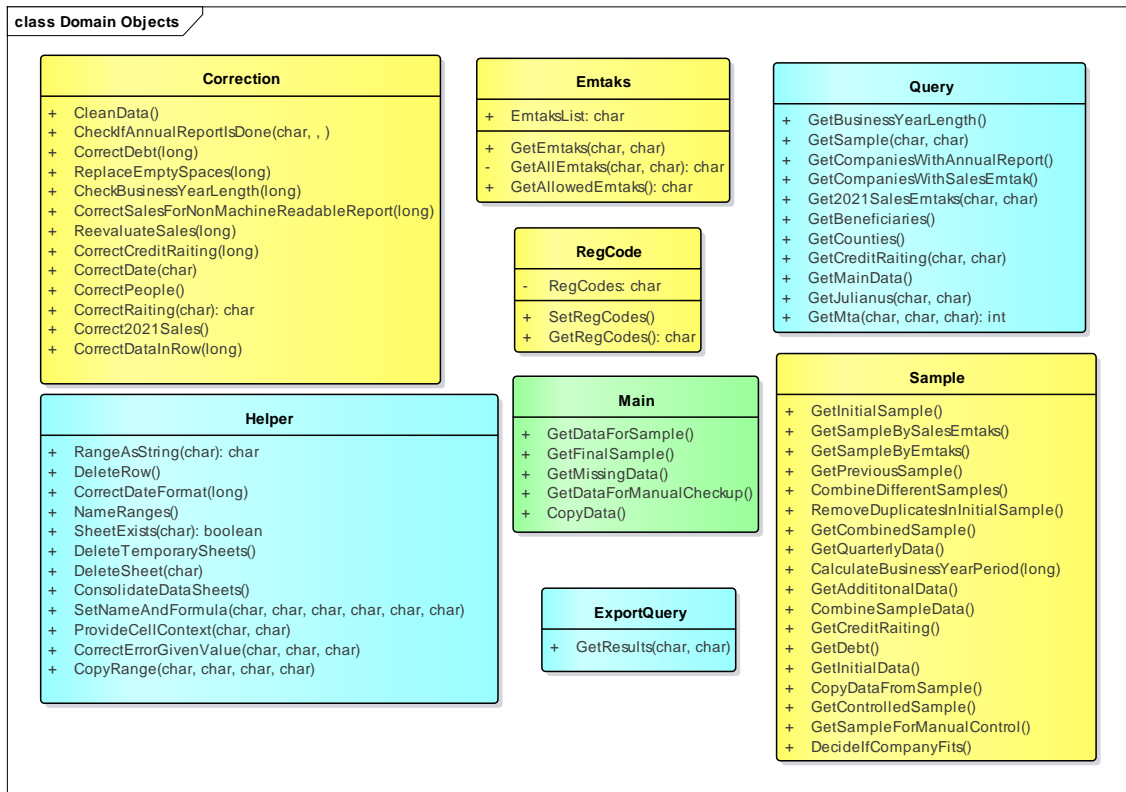
- Programm peab saama kõik andmed kätte andmebaasist ilma, et kasutaja peaks tegema manuaalseid andmebaasipäringuid..
- Programmi sisendiks on eelmine valim, sobilikud ja sobimatud EMTAK-id ning käibepiir. Eelmine valim, sobivad EMTSK-id võimaldavad saada maksimaalse esialgse valimi, et sealt filtreerida välja ettevõtted mis valimisse ei sobi. Välja filtreerimiseks on oluline käibepiir ja samuti EMTAK-id, et näha milline ettevõtte ei ületa käibepiiri või millise ettevõtte majandusaasta aruandes “Müügitulu jaotus tegevusalade lõikes” osas märgitud põhitegevusala EMTAK ei ole etteantud sobilike EMTAK-ide osas.
- Programm peab tuvastama antud EMTAK-i tütar-EMTAK-id, selleks et kasutaja ei peaks sisestama kõik EMTAK-id käsitsi.

- Esialgse valimi failis peab olema ära toodud ettevõtte kohta järgnevad asjad:
  - registrikood ja ärinimi, mis on vajalikud ettevõtte identifitseerimiseks
  - registreeritud EMTAK ja selle kirjeldus, et näha mis valdkonna ettevõtte on ettevõtte registreeritud
  - 2021.a müügitulu ja töötajate arv (majandusaasta aruandest). Müügitulu on üks valimisse saamise sisenditest. Töötajate arvu kasutatakse selleks, et kontrollida, kas mõne ettevõtte müügitulu on puudu ja see tuleb käsitsi sisestada või kas aruanne on esitatud ning selle tulemused on olemas, kuid deklareeritud müügitulu on 0. Selleks kasutatakse just 2021. aasta majandusaasta aruande tulemusi, sest see on kõige hilisem aasta, mille tulemused peavad olema juba esitatud.
  - 2022.a 2.kvartali kuni 2023.a 1.kvartali käive (MTA deklaratsioonist). Valimisse saamiseks on vaja arvutada viimase nelja kvartali koondkäive, mis saadakse 2022. aasta teise kvartali kuni 2023. aasta esimese kvartali käibe deklaratsioonidest. Kui koondkäive ületab käibepiiri, võib ettevõtte siiski valimisse saada, kui ta vastab teistele tingimustele.
  - esialgses valimis olemise põhjus, see aitab paika panna millised on valimisse saamise võimalused antud ettevõttel.
  - Majandusaasta aruande põhitegevusala EMTAK ja selle kirjeldus, sest need on aluseks ettevõtte tegevusala määramisel. Majandusaasta aruande jaotises "Müügitulu jaotus tegevusalade lõikes" on välja toodud ettevõtte peamine tegevusala majandusaasta aruandes käsitletaval aastal, mis on kõige uuem koht, kust saada infot ettevõtte tegevusala kohta. Tihtipeale ei muudeta registreeritud EMTAK-koodi peale tegemisala vahetamist
  - kas on lubatud EMTAK koodide hulgas
  - kas klappib registreeritud EMTAK-iga
  - otsus kas jääb valimist välja, läheb valimisse või vajab käsitsi üle kontrollimist

- kas viimase nelja kvartali müügitulu summa on vähemalt 75% käibepiirist, kui nende summeeritud kvartaalne käive ületab 75% piiri, siis on tõenäosus, et nende 2021.a majandusaasta aruandes olev müügitulu ületab käibepiiri, ning esialgses valimis olevat tulemust, on vaja korrigeerida.
- Peale esialgse valimi koostamist peab olema võimalik ajakohastada majandustulemusi ettevõtetel, kelle majandusaasta aruanne on masinloetamatud, et kõiki ettevõtteid saaks võrrelda samadel tingimustel olenemata majandusaasta aruande esitamise vormist. ning kelle summeeritud kvartaalne käive ületab vähemalt 75% käibepiirist.
- Peale esialgse valimi üle vaatamist tuleb valimisse sobivad ettevõtted *export*-ida teisele lehele, et valimi ettevõtteid üle vaadates ei peaks neid filtreerima suurest valimist.
- Valimisse jäänud ettevõtete jaoks tuleb importida järgnevad asjad:
  - Müügitulu ja töötajate arv 2020.a ning 2021.a majandusaasta aruandest, et näidata milline on olnud ettevõtte suurus (töötajate arvu järgi) viimastel aastatel ja milline on olnud nende majandustulemus (müügitulu)
  - 2020 2.kv - 2023 1.kv käive, tööjõumaksud ja töötajate arv (MTA), need on majandustulemused, mille põhjalt hakatakse looma ettevõtete paremusjärjekordasid
  - MTA ja inkasso võlg seisuga 01.02.2023 ja 01.03.2023, et näidata millised on ettevõtte makseharjumused
  - Krediidihinnang seisuga 11.03.2023 ja 11.04.2023, et näidata Äripäev Infopanga poolt välja kalkuleeritud hinnangut ettevõttele võttes arvesse ettevõtte suurust, vanust ja makse harjumusi
  - Infopanga kasusaajate list, et näidata kes saavad kasu ettevõtte majandustegevustest
- Andmete importimine peab olema kasutajasõbralik, lihtne ja intuitiivne, et vältida kasutajate segadust või vigu.

- Tarkvara peab olema dokumenteeritud, nii et kasutajad saaksid vajadusel kergesti juhendit järgida või tehnilist tuge saada.

### 3.2 Arhitektuur ja disain



Joonis 2. Ülevaade rakenduses kasutatavatest moodulitest.

Autor lõi rakenduse tegemiseks kaheksa erinevat moodulit. Koodi moodulite vahel jagamine võimaldab rakenduse arendamisel ja haldamisel paremini orienteeruda. Joonisel 2 on kujutatud kõik moodulid vastavalt nende rakendusele. Rohelisega on peamine moodul, mis täidab nõu “Main” klassi ülesannet. Kollasega on märgitud moodulid, mis sisaldavad peamist ärilist loogikat ja mille funktsioonid kutsutakse välja “Main” modelis. Moodulid on jaotatud neljaks. Korduvalt kasutatavad registrikoodiga seotud asjad on moodulis “RegCode” ning EMTAK-idega seotud loogika on moodulis “Emtaks”. Kogu äriline loogika on pandud moodulisse “Sample”. Moodulis “Correction” on kogu loogika andmete puhastamiseks. Äriliste funktsioonide toetamiseks on loodud toetavad moodulid. Moodulid “Query” ja “QueryExport” käivad käsikäes. Moodulis “Query” on välja toodud andmebaasipäringud, mille tulemusi salvestatakse Excelisse mooduli “ExportQuery” funktsiooni abil. Lisaks on moodul “Helper”, mis sisaldab erinevaid



abistavaid funktsiooni, mis ei käi konkreetse mooduli või loogika osa juurde. Seal on meetodid näiteks Exceli lehekülgede või ridade kustutamiseks.

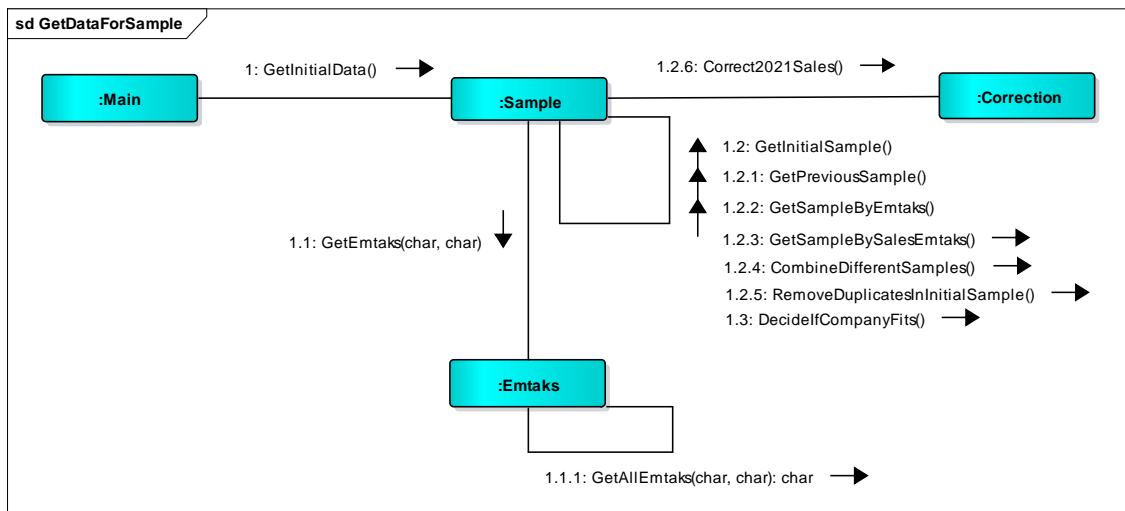
Autor kasutas mooduleid, et luua tarkvaralahendus, mis vastavalt varasemalt paika pandud nõuetele jaotub viieks eraldi käivitatavaks osaks. Esimeses etapis saadakse vajalikud andmed, et luua esialgne valim. Teises etapis valitakse välja ettevõtted, mis lähevad manuaalselt kontrolli ning kolmandas etapis luuakse lõplik valim vastavalt manuaalse kontrolli tulemustele. Neljandas etapis saadakse juurde kõik vajalik info, mida varem ei olnud. Viimases ehk viiendas etapis liidetakse kõik info kokku üheks suureks andmetabeliks.

### **3.2.1 Esialgse valimi koostamine**

See VBA kood on mõeldud andmete tagastamiseks Exceli töövihiku "Input" töölehe kaudu ja ettevõtete valimiseks kindlate kriteeriumite alusel. Valimi loomiseks kasutatakse erinevaid alamprogramme, mis tagastavad andmeid EMTAK-koodide, varasema valimi andmete ja müügiga seotud EMTAK-koodide alusel. Esialgsest valimist eemaldatakse duplikaadid ja defineeritakse nimed kasutatavatele Exceli lahtrite vahemikele, et neid oleks hiljem valemite lihtsam kasutada.

Selle meetodi sisenditeks on vajalikud sobivate EMTAK-koodide nimekiri Exceli lehel "Input" lahtris "E1" ning mittesobivate EMTAK-koodide nimekiri lahtris "E2". Lisaks on vajalik käibepiiri väärtus, mis sisestatakse lahtrisse "E3". Meetodi GetPreviousSample sisendiks on eelmise valimi ettevõtete registrikoodid, mis asuvad tulbas "A". Eeldatakse, et on olemas lehed "Input" ja "InitialSample" ning ei ole olemas lehti "AnnualReport", "SampleBySales", "SampleByEmtaks", "EligibleAccordingToSales".

Selle protsessi tulemusena luuakse esialgne valim, mis kuvatakse lehel "InitialSample". Seal kuvatakse igale ettevõttele vastavalt andmebaasi päringutele registrikood, nimi, EMTAK kood ja selle kirjeldus, ettevõtte ID, müügitulu 2021. aastal, töötajate arv ja MTA poolt deklareeritud käive 2022. aasta teisest kvartalist kuni 2023. aasta esimese kvartalini. Lisaks sellele kuvatakse iga ettevõtte müügitulu jaotusele vastav EMTAK ja selle kirjeldusele, mis on saadaval ka lehel "SalesEmtak". Lehel on ka märgitud, kuidas ettevõtte valimisse sattus, kas selle müügitulu EMTAK vastab sobivatele EMTAK-idele ning kas ettevõtte sobib valimisse. Lisaks on veel loodud lehekülg "AnnualReport".



Joonis 3. Suhtlusdiagramm esimese valimi koostamise jaoks.

Oodatud tulemuse saavutamiseks tehtavad sammud on näha joonisel 3. Joonise loetavuse säilitamiseks on jäetud kajastamata toetavad moodulid “Helper”, “Query” ja “QueryExport”.

Kõigepealt fikseeritakse ära EMTAK-id, mis on aktsepteeritud. Peale nende paika panemist luuakse vastavalt EMTAK-idele ja eelmisele valimile uus esialgne valim. Kuna kõikidel ettevõtetel ei ole majandusaasta aruanne masinloetav, siis nende müügitulu väärtused on tabelist puudu. Selle parandamiseks kontrollitakse, kas ettevõtte on majandusaasta aruande esitanud ja need, kes ei ole, nendele pannakse müügitulu kohapeale 0. Sisestatakse ka valemid, mis hindavad ettevõtte sobivust valimisse.

Selleks, et faili andmemahat ei muutus liialt suureks ja andmeid ei oleks dubleeritud kustutatakse ära üleliigsed ajutised leheküljed. Selleks, et allesjäänud olulist infot oleks edaspidi lihtsam kasutada, pannakse erinevatele tabelitele ja nende tulpadele kirjeldavad nimed.

Vastavalt esialgse valimi suurusele ning EMTAK-koode arvule ja populaarsusele võib koodi tööaeg varieeruda.

### 3.2.2 Manuaalne kontroll

Makro "GetDataForManualCheckup" kopeerib valitud andmed töölehel "InitialSample" ning kleebib need töölehele "ManualCheckup". Seejärel rakendab ta automaatfiltrid, et kuvada ainult "control" kriteeriumile vastavad read ehk read, mis vajavad manuaalselt kontrollimist.

Andmed, mida kopeeritakse, asuvad aktiivsel töölehel "InitialSample" tabelis "Table1". Kood eeldab, et töölehed "ManualCheckup" ja "InitialSample" on juba olemas.

Makro kopeerib valitud andmed töölehele "ManualCheckup" lahtrisse "C1". Seejärel rakendatakse automaatfiltrid, et kuvada ainult read, kus veeru 6 väärtus on "control". Lõpuks kustutatakse veerud "E:M" ja "I:J".

### **3.2.3 Lõpliku valimi koostamine**

Makro koosneb kolmest alamprotseduurist: "FilterTable", "CreateDocuments" ja "GetFinalSample". Makro eesmärk on filtreerida andmeid Exceli tabelist "ManualCheckup" ja luua uus tööleht "Sample", kuhu jäetakse alles kõik ettevõtted, mis on kas peale esmast kontrolli või manuaalset kontrolli otsustatud valimisse võtta.

Makro sisenditeks on Exceli tabeli "ManualCheckup" andmed, mida töödeldakse. Andmed on paigutatud tulpadesse C:I, kusjuures andmete esimene rida sisaldab väljade nimesid (nt "Field").

Makro väljundiks on uus tööleht "Sample" koos filtreeritud andmetega. Andmed on paigutatud tulpadest A kuni G. Programm on loonud ja kustutanud ajutise lehe "Temp".

Lõpliku valimi koostamise sammud on näha joonisel 4. Makro "CreateDocuments" kopeerib valimi jaoks sobilikud ettevõtted eraldi Exceli töövihikulehele, mida kasutatakse edasise analüüsi jaoks. Peale seda luuakse ajutine lehekülg "Temp", kus valitakse õigete filtreerimiskriteeriumiga read ja kopeeritakse lehele "Sample". Seejärel kustutatakse "Temp" lehekülg.

Makro kasutab Exceli funktsioone nagu "AutoFilter", "Copy", "PasteSpecial" ja "Delete", et automatiseerida andmete töötlemist ja töölehtede loomist Exceli keskkonnas.

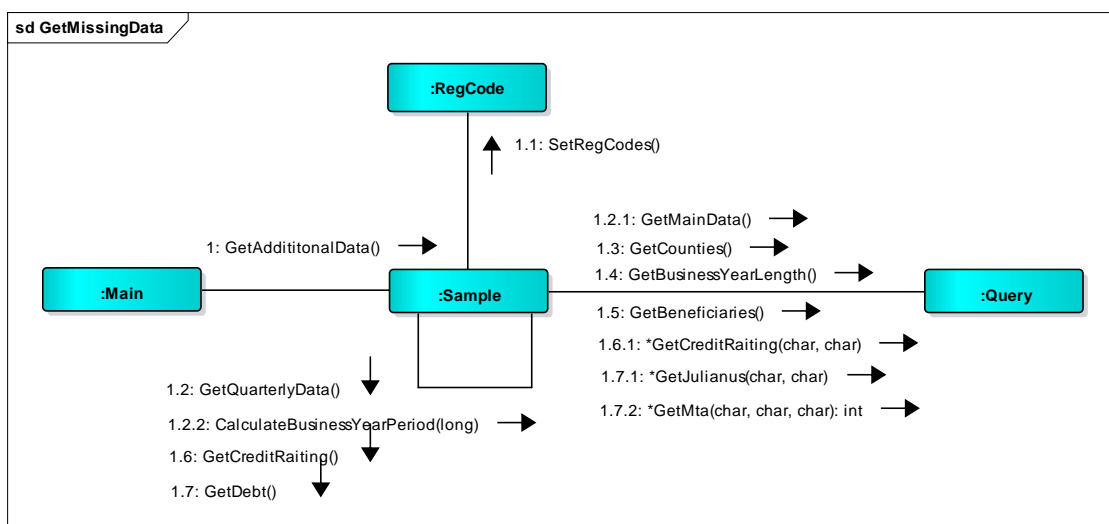
### **3.2.4 Andmete kogumine**

Makro eesmärk on kasutada ettevõtete registrikoodide andmeid lehelt "Sample" ning erinevaid andmebaasipäringuid, et saada täiendavaid andmeid andmebaasist. Need täiendavad andmed on vajalikud raportite sisendiks oleva andmefaili loomiseks. Kõik päringute tulemused salvestatakse eraldi lehtedele.

Esimeseks sisendiks on leht "Sample" ja seal olevate ettevõtete registrikoodid. Makro sisendiks on ka teksti kujul andmebaasipäringud maakondade, majandusaasta pikkuse, kasusaajate, krediidihinnangu, MTA võlgade ning inkasso võlgade leidmiseks. Sisendiks on veel krediidihinnangu ning võlgade jaoks kuupäevad, mis seisuga infot on vaja.

Eeldatakse, et lehed "MainData", "Counties", "BusinessYearLength", "Beneficiaries", "CreditRatingOlder", "CreditRatingNewer", "JulianusOlder", "JulianusNewer", "MtaOlder" ja "MtaNewer" ei ole veel loodud.

Koodi väljundid on mitmed Exceli töölehed, kuhu kood ekspordib erinevat tüüpi andmeid. Näiteks on olemas töölehed "MainData", "Counties", "BusinessYearLength", "Beneficiaries", "CreditRatingOlder", "CreditRatingNewer", "JulianusOlder", "JulianusNewer", "MtaOlder" ja "MtaNewer". Iga tööleht sisaldab vastavate subruutiniide poolt tagastatud andmeid.



Joonis 4. Jadadiagramm andmete kogumise sammude näitamiseks.

Makro poolt läbitavad sammud on näha joonisel 4. Jooniselt on välja jäetud toetavad moodulid "Helper" ja "QueryExport".

Makro koosneb mitmest alamprogrammist, mis taastavad andmeid SQL andmebaasist ja ekspordivad need Exceli. Põhialamprogramm on nimega "GetAdditionalData" ning see kutsub välja mitmeid teisi alamprogramme, mis igaüks pärivad erinevat tüüpi andmeid. Koodi eesmärk on automatiseerida andmete importimise protsessi, muutes selle kiiremaks ja tõhusamaks.

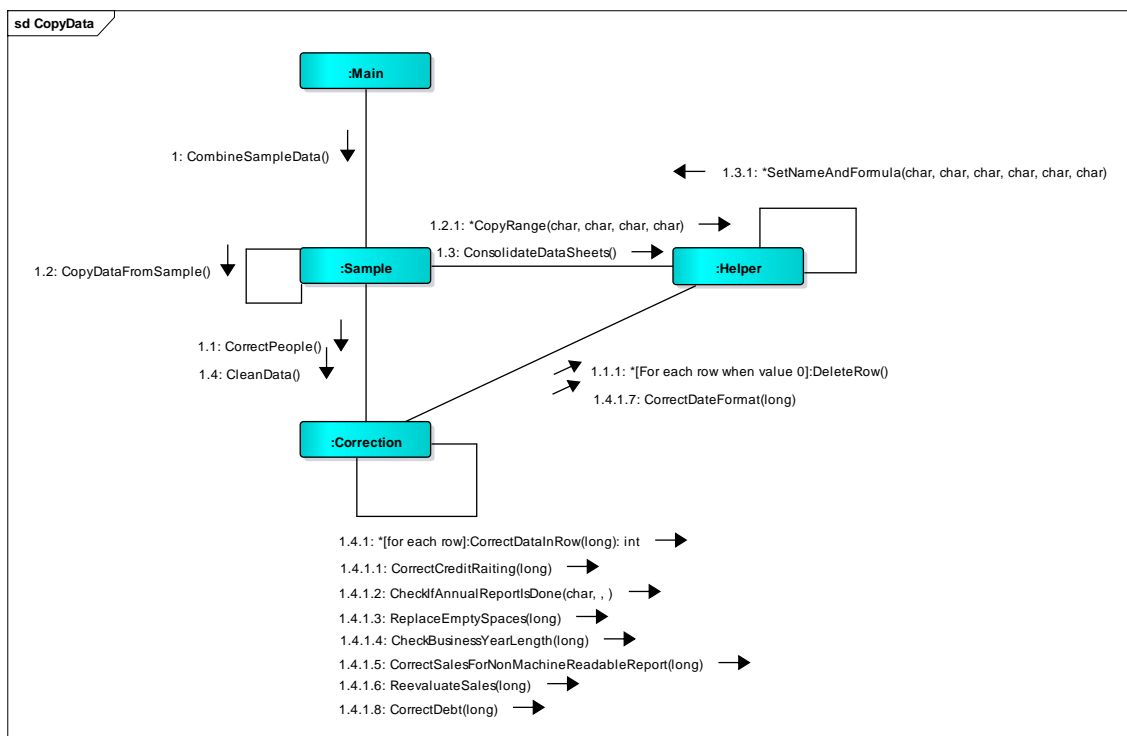
Kõigepealt pannakse paika millised on valimis olevate ettevõtete registrikoodid. Kui on teada milliseid ettevõtteid otsitakse, siis tehakse päringud MTA-st saadud kvartaalsete andmete saamiseks. Peale seda tehakse päringud maakondade, majandusaasta pikkuste, ja kasusaajate leidmiseks. Lõpetuseks kontrollitakse kas erinevate kuupäevade seisudega on ettevõtetel olnud inkasso või MTA võlgasid.

Kuna koodi jooksmise aeg varieerub vastavalt valimi suurusele, siis antakse kasutajale sõnumi kastikese abil teada kui päringud on valmis saanud.

### 3.2.5 Andmetabeli koostamine

Makro eesmärk on kopeerida andmed erinevatelt vahelehekülgedelt kokku leheküljele “Data” ja puhastada kopeeritud andmed.

Selle meetodi sisendiks on erinevad leheküljed, kuhu punktis 4.2.4 andmed salvestati ning meetodi väljundiks on andmetabel leheküljel “Data”.



Joonis 5. Suhtlusdiagramm andmetabeli koostamise makrost.

Andmete kopeerimiseks ühte tabelisse kasutatakse meetodit “CopyData”. Enen kui andmeid saab kopeerima hakata, on vaja teha natuke korrekture. Selleks, et vältida võimalust et mõni ettevõte, mis ei peaks olema valimis jääb lõppfaili, tühjendatakse esimese asjana tabel. Selle järel korrigeeritakse kasusaajate kajastamist tabelis.

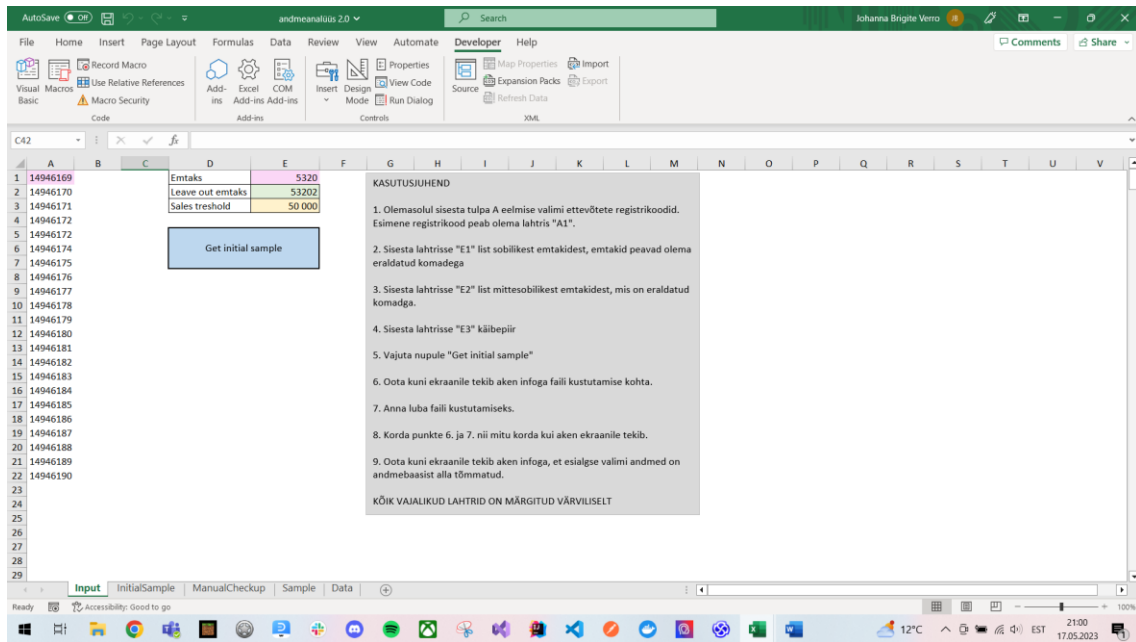
Andmebaasist saadud väärtuste puhul on iga kasusaaja eraldi real. Tabelis on vaja aga kuvada iga ettevõtte kasusaajaid listina.

Peale seda on kõikides failides andmed õigel kujul. Järgmise asjana kopeeritakse kõik andmed kokku ühte tabelisse lehel "Data". Pärast andmete kokku kopeerimist ühte tabelisse puhastatakse andmed. Korrigeeritakse krediidi hinnanguid vastavalt uuele standardile, lisatakse igale ettevõttele märge, kas majandusaasta aruanne on esitatud, milline oli majandusaasta pikkus ja kas maksu- või inkassovõlad on olemas. Lisaks arvutatakse ümber müügitulu, kui majandusaasta pikkus ei olnud 12 kuud, ning standardiseeritakse registreerimiskuupäeva esitamise tabelis. Meetod täidab ka kommentaaride lahtrit vastavate märkustega, näiteks kui müügitulu on korrigeeritud või kui majandusaasta pikkus erineb.

Vastavalt valimi suurusele võib koodi tööaeg varieeruda.

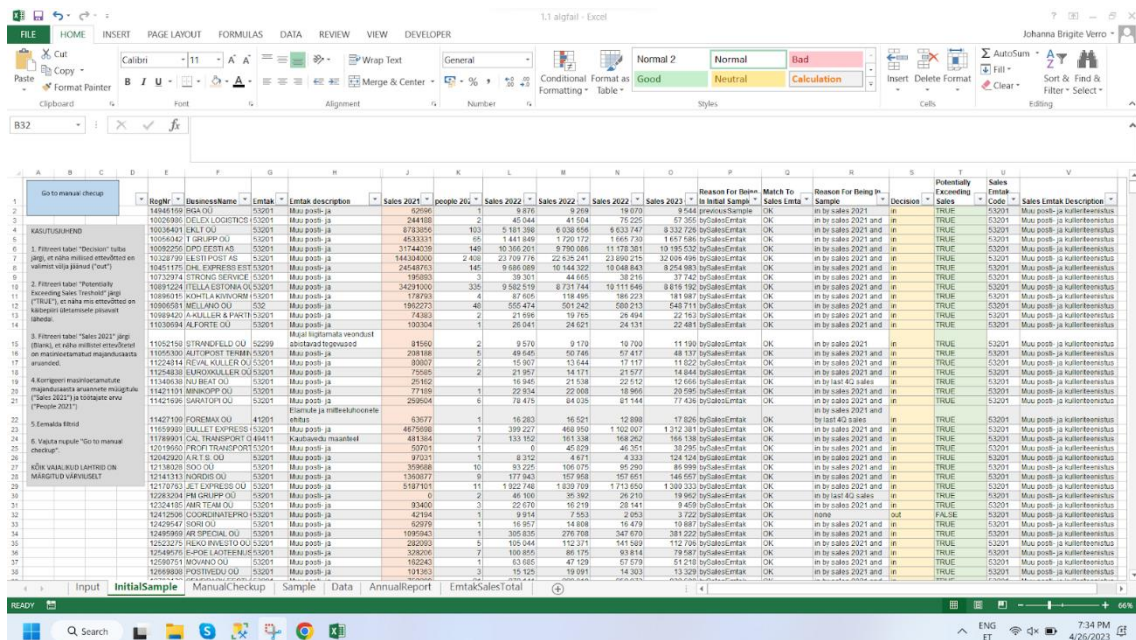
### **3.3 Kasutajaliides**

Iga lehele on lisatud kasutusjuhend, mis aitab kasutajal programmi lihtsamini navigeerida. Olulised lahtrid ja tulbad on märgitud erinevate värvidega, et need oleksid kasutajale paremini nähtavad. Selleks, et kasutajal oleks lihtsam erinevate programmide vahel orienteeruda lõpeb iga programm lehel, kus tuleb teha järgmised liigutused andmetabeli valmis saamiseks.



Joonis 6. Kasutajaliides lehel "Input".

Esialgse valimi saamiseks peab kasutaja sisestama sobivad ja mittesobivad EMTAK-id lehele "Input" vastavatesse lahtritesse "E1" ja "E2". Lisaks tuleb määrata käibepiir väärtuses "E3" ning sisestada eelmise valimi ettevõtete registrikoodid tulpadesse "A". Seejärel saab kasutaja vajutada nupule "Get initial sample", et rakendada meetodit "GetDataForInitialSample", mis on kirjeldatud punktis 4.2.1.



Joonis 7. Kasutajaliidese vaade lehel "InitialSample".

Näha on tabelit esialgse valimiga. Kasutaja saab filtreerida tabeli veergude "Sales 2021", "Decision" ja "Potentially Exceeding Sales" alusel, et näha ettevõtteid, kelle majandusaasta müügitulu vajab korrigeerimist. Pärast tulemuste korrigeerimist saab kasutaja filtrid eemaldada ning vajutada nupule "Go to manual checkup", mis käivitab meetodi "GetDataForManualCheckup", nagu kirjeldatud punktis 4.2.2, et minna käsitsi kontrollima ettevõtteid, kelle majandusaasta aruanded tuleb üle vaadata.

	RegNr	BusinessName	Reason For Being In	Initial Sample	Match To Sales Emtak	Sample	Decision	Field
110	11442250	ROADRUNNER OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
145	11389975	FEDERAL EXPRESS CORPORATION EESTI FILIAAL	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
149	12245296	AIRPACK ESTONIA OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
153	12768344	PAKIPPOINT OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
184	12616479	BENTITE OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
239	16020298	NI TRANSPORT OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
248	16317779	ALMRK OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
250	16352226	KULLERON OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
292	16338741	LAKKOS OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
294	16453531	BIOSENSE OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
306	14061092	KARVEL OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
316	14713488	SETESENG DIHLABAKELA OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
318	14823863	VELOVAN OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
329	16296365	HAYFARMA OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
330	16428504	READYSTEADYGO OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
331	16496612	LINCOLN OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
332	16517389	KOXXON OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
333	16539238	LINEOFF GROUP OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
336	16597253	TREE DELIVERY OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
343	16334864	OVOS OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
351	16480545	TODOOR OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
367	16338563	VOVS OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
369	16363785	BENCH OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
370	16374211	FLEETY OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
371	16422111	DELOVERTY OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		
372	16443952	FOONT OÜ	byRegisteredEmtak	sales by emtak is missing	in by last 4 quarter sales	control		

Joonis 8. Kasutajaliides lehel "ManualCheckup".

Tabelis kuvatakse ettevõtteid, mille majandusaasta aruandeid tuleb käsitsi üle vaadata, et otsustada, kas ettevõtte kuulub valimisse või mitte. Kasutaja saab kontrollida ettevõtteid Infopanga lehel või muul lehel ning seejärel langetada otsuse, sisestades tulpa "Decision" "in", kui ettevõtte tuleb valimisse võtta, või "out", kui ettevõtte tuleb välja jätta. Kui raportis on korrara rohkem kui üks sektor, saab tulpa "Field" sisestada vastava sektori nime. Pärast ettevõtteid kontrollimist saab kasutaja eemaldada filtrid ja luua lõpliku valimi, vajutades nupule "Create sample".



RegNr	BusinessName	Reason For Being In Initial Sample	Match To Sales	Reason For Being In Sample	Decision	Field
14946169	BGA OÜ	previousSample	OK	in by sales 2021	in	Muu posti- ja kulleriteenistus
10026986	DELIX LOGISTICS OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10038401	ENLET OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10086042	T GRUUP OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10092256	DPD EESTI AS	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10328799	EESTI POST AS	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10451175	DHL EXPRESS ESTONIA AS	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10732974	STRONG SERVICE OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10891224	ITELLA ESTONIA OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10896015	KOHTLA KIVIVORM OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10906581	MELLANO OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
10989420	A KULLER & PARTNERID OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11030694	ALFORTE OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11052158	STRANDELD OÜ	bySalesEntak	OK	in by sales 2021	in	Muu posti- ja kulleriteenistus
11055300	AUTOPOST TERMINAL OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11224814	REVAL KULLER OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11254838	EUROXKULLER OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11340638	NU BEAT OÜ	bySalesEntak	OK	in by last 4Q sales	in	Muu posti- ja kulleriteenistus
11421101	MINIKOPP OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11421696	SARATOPI OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11427109	FOREMAX OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11659989	BULLET EXPRESS OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
11789901	CAL TRANSPORT OÜ	bySalesEntak	OK	in by sales 2021 and by last 4Q sales	in	Muu posti- ja kulleriteenistus
12019660	PROFI TRANSPORT OÜ	hwSalesEntak	OK	in hw sales 2021 and hw last 4Q sales	in	Muu posti- ja kulleriteenistus

Joonis 9. Kasutajaliides lehel “Sample”.

Lehel “Sample” on näha tabelit lõpliku valimiga. Kasutajal on võimalik valimist ettevõtteid eemaldada või lisada. Kui valim on soovi korral üle vaadatud saab kasutaja vajutada nupule “Get data for analysis”. See nupp käivitab meetodi “GetAdditionalData”, mis laeb alla kogu rapordi jaoks vajaliku info, mida pole veel andmebaasist alla tõmmatud.

Status	Kommentaar	Valdkond	Majandusaasta	Aruanne	Majandusaasta pilku	Maakond	Asutamisaeg	Regkood
KASUTUSÜHEND								
1. Vajuta nupule "Get data for report".								
2. Oota kuni ekraanile ilmub aken infoga, et andmed saidi kopeeritud								

Joonis 10. Kasutajaliidese vaade lehel “Data”.

Lehel "Data" on näha lõplikku andmetabelit, mida taheti tarkvaraga saavutada. Samuti on seal nupp "Get data for raport", mis käivitab meetodi "CopyData" (välja toodud punktis 4.2.5), mis kopeerib erinevatelt lehtedelt vastavad andmetabeli väärtused.

## **4 Analüüs ja järeldused**

### **4.1 Tehnoloogia valik**

Autor otsustas konkurentsiraportite analüüsi automatiseerimisel jätta analüüsi põhjaks Exceli. Valiku juures sai määravaks mitu aspekti. Esiteks, varasem lahendus põhines Excelil. Teiseks, ettevõtte poolne nõue oli, et tarkvaralahendus peab töötama kõikides arvutites. Seda arvestades võib oletada, et andmeanalüüsi võivad kasutada erinevad inimesed, kellel on ka erinevad oskused. Varasem kirjanduslik materjal andis autorile kinnitust, et Excel on algajatele kasulik ning sobilik tööriist. Veel üks Exceli eelis on see, et kõikidel töötajatel ei pruugi olla muid sobivaid analüüsiks mõeldud tarkvaralahendusi. Seega otsustas autor jääda Exceli juurde ning valida sobiva programmeerimiskeele, millega analüüsi automatiseerida.

Autor leidis vastavalt teiste varasemate tööde ja uuringute tulemustele kolm sobivat programmeerimiskeelt: Python, R ja VBA. Need keeled erinevad oluliselt teineteisest ja neil on erinevad tugevused ja nõrkused. VBA on spetsiaalselt loodud Microsoft Office'i programmide automatiseerimiseks, kuid selles on vähem erinevaid andmestruktuure kui Pythonis või R-is. R tugevuseks on spetsialiseeritus statistikale ning populaarsus andmekaevanduses ja andmeanalüüsis. Üks R-i miinuseid on, et selles ei ole nii tugevaid visuaalse andmete esitamise võimalusi kui Pythonis. Pythonil on jällegi lai valik andmestruktuure, kuid selle rakendamine pole alati nii lihtne kui VBA.

Autori valik jäi peamiselt Pythoni ja VBA vahele. VBA-le andis lõpuks ülekaalu see, et tegu on siiski spetsiaalselt Microsoft Office-is kasutamiseks mõeldud keelega ja selle rakendamine on Excelis selle võrra lihtsam. Otsusele aitas kaasa ka see, et VBA on ettevõtte siseselt juba kasutusel ehk selle juurutusprotsess konkurentsiraportite puhul oleks lühem.

### **4.2 Efektiivsus ja täpsus**

Autor otsustas valideerida tulemusi, võrreldes erinevate näitajate põhjal, kui kaua läheb aega andmefaili koostamisel käsitsi ja rakendust kasutades. Käsitsi meetodil, kasutades varasemat süsteemi, möödab autor aega andmebaasipäringute tegemiseks, sisendite käsitsi sisestamiseks päringute jaoks, ning päringu tulemuste salvestamiseks Exceli

lehtedele. Seejärel sisestatakse valemite abil tulemused ühte Exceli andmetabelisse, mida kasutatakse edasise andmeanalüüsi sisendina. Autori loodud rakenduse puhul on vastandiks punktis 4.2.1 välja toodud makrole sisendite andmine, esialgse ning lõpliku valimi koostamine, lisa andmete pärimine ning lõpliku andmetabeli koostamine.

Autor otsustas jätta võrdlusest välja aja, mis kulub ettevõtete käsitsi ülevaatomiseks, kuna loodud tarkvara valib ettevõtteid kolmel erineval tingimusel, samas kui varasem süsteem võttis arvesse ainult ühe tingimuse. Seetõttu võib uuel automatiseeritud versioonil olla rohkem potentsiaalseid kandidaate, keda on vaja käsitsi üle vaadata, mis muudab võrdluse ebavõrdseks.

Valimik koostamise aeg sõltub suuresti sellest, kui paljusid valdkondi on vaja ühes valimis hõlmata. Erinevatel sektoritel võib olla erinev käibepiir valdkondade kaupa, mis võib aeglustada valimi koostamist ja lisada täiendavat tööd. Autor kontrollis süsteemi tulemusi erinevate valdkondade arvuga sektorites, et testida erinevaid realistlikke stsenaariume. Kui valimis oli mitu valdkonda, siis autor ei arvestanud uute ettevõtete valdkondadesse jaotamiseks kulunud aega. Ajakulu hõlmas ainult aega, mis kulus varasemalt valimisse kuulunud ettevõtete valdkondade määramiseks.

Esialgses valimis olevate ettevõtete arv	Valdkondade arv	Vana süsteemi aeg	Uue valimi ettevõtete arv vana süsteemi järgi	Uue süsteemi aeg	Uue valimi ettevõtete arv uue süsteemi järgi
330	1	1h 34min 16sek	410	5min 54sek	524
245	2	1h 26min 39sek	482	8min 32sek	857
924	3	2h 42min 06 sek	768	15min 12sek	924
262	5	1h 47min 35sek	239	7min 18sek	318

Joonis 11. Ülevaade efektiivsus testide tulemustest.

Autor viis läbi efektiivsuse testi, et võrrelda automatiseeritud ja käsitsi valimi koostamise meetodeid erinevate valdkondade arvuga sektorites. Testi esimene osa hõlmas ühe valdkonnaga sektorit, kus algseks valimiks oli 330 ettevõtet. Tulemused näitasid, et automatiseeritud meetod oli oluliselt kiirem kui käsitsi meetod. Andmete analüüsimine võttis automatiseeritud meetodil aega 6 minutit, samas kui käsitsi meetod võttis aega umbes 1,5 tundi. Seega oli automatiseeritud meetod umbes 15 korda kiirem kui käsitsi meetod.

Valimi suurus erines kahe meetodi vahel. Käsitsi meetodi puhul jäi lõppvalimisse 410 ettevõtet, mis tähendas valimi suuruse muutust 24%. Samas oli automatiseeritud meetodi puhul lõppvalimis 524 ettevõtet, mis tähendas 51,8% suurust valimi kasvu. On oluline märkida, et käibepiiri ei korrigeeritud, seega võib suurem valimi kasv olla osaliselt tingitud kasvavatest hindadest. See tähendab, et suurem hulk ettevõtteid võib ületada käibepiiri.

Autor viis läbi ka testi kahe, kolme ja viie valdkonnaga sektoritega. Testi tulemused näitasid, et automatiseeritud meetod oli ka nendes sektorites kiirem kui käsitsi meetod. Kahe valdkonnaga sektorites võttis automatiseeritud meetod lõppvalimi koostamiseks aega 8 minutit ja 32 sekundit, samas kui käsitsi meetod võttis aega 1 tund 26 minutit ja 39 sekundit. Viie valdkonnaga sektorites võttis aega vastavalt 7 minutit ja 18 sekundit ja 1 tund 47 minutit ning 35 sekundit.

Kolme valdkonnaga sektoris oli seis natuke keerulisem. Seal kasutati varianti, kus erinevatel valdkondadel oli erinev käibepiir. See tähendas, et valimi koostamine oli vaja jagada kolme osasse, vastavalt valdkonnale, selleks et igal valdkonnal saaks kasutada vastavat käibepiiri. See oli olukord, kus autori loodud rakendust oli vaja jooksutada kolm korda. Tänu sellele võttis automatiseeritud süsteem aega 15 minutit ja 12 sekundit, kui käsitsi meetod võttis aega märkimisväärselt rohkem - 2 tund 42 minutit ja 6 sekundit.

Kokkuvõtlikult võib öelda, et vana lahenduse järgi keskmiselt ühe raporti andmefaili koostamine võtab aega 1 tund 52 minutit ja 14 sekundit, samas kui autor loodud lahendust kasutades on keskmine aeg ainult 9 minutit ja 39 sekundit. Seega on uus lahendus vana lahendusega võrreldes ligi 12 korda kiirem ning võimaldab efektiivsemat töövoogu andmefailide koostamisel.

### **4.3 Põhjendus**

Tuginedes tulemustele eelmises punktis võib öelda, et automatiseeritud meetod oli kiirem ja otstarbekam kui käsitsi meetod. Seetõttu soovitab autor kasutada automatiseeritud meetodit valimi koostamisel ja andmete analüüsimisel erinevate valdkondade arvuga sektorites. Seda soovitus toetab fakt, et andmete analüüsimise tõhusus ja kiirus on märkimisväärselt paranenud, võimaldades tulemuste kiiret ja täpset kättesaadavust. See

annab ettevõttele konkurentsieelise, sest võimaldab raporteid koostada tunduvalt kiiremini, mis on selles valdkonnas üks olulisemaid tegureid.

On oluline, et raportid oleksid väljas võimalikult kiiresti, kuna viivitused võivad tähendada, et teised ettevõtted jõuavad turule enne seda, kui Infopanga toode on valmis. Seega, hoolimata sellest, et Infopanga raportid võivad olla paremad, kui need avaldatakse mitu nädalat pärast konkurentide omi, võib see negatiivselt mõjutada toote müüki. Seetõttu on kiirus ja efektiivsus selle valdkonna jaoks kriitilise tähtsusega ning automatiseeritud versiooni kasutamine võib aidata tagada, et raportid valmiks võimalikult kiiresti ja täpselt.

Selle lõputöö tulemused näitavad, et automatiseeritud versioon võtab andmete analüüsimiseks oluliselt vähem aega kui käsitsi tehtav versioon. Tulemuste põhjal võib järeldada, et automatiseeritud protsess on palju efektiivsem ja aja kokkuhoiuga seotud eelised on märkimisväärsed. Seega on soovitus automatiseeritud meetodi kasutamiseks laialdaselt põhjendatud.

#### **4.4 Kitsaskohad ja edasiarendused**

Kuigi tarkvara kasutamise õigustatus ja selle plussid võivad olla tõestatud, on siiski oluline märkida, et igal tarkvaral on oma piirangud ja kitsaskohad. Selle tarkvaralahenduse üks kitsaskohti seisneb selles, et suurte andmemahtude töötlemine võib võtta palju aega, eriti kui kasutatakse VBA-d. VBA korral tuleb suur andmemaht salvestada Exceli faili ja seejärel uuesti lugeda, mis võtab aega ja vähendab analüüsi efektiivsust. Üheks lahenduseks oleks kasutada mõnda teist programmeerimiskeelt, nagu Python või R, mis pakuvad paremat andmete salvestamise võimalust ja laiemat valikut erinevaid andmestruktuure.

Python ja R võimaldavad luua samasuguseid funktsioone ja makrosid nagu VBA-s, kuid andmete salvestamise võimalused on paremad. Näiteks võib Pythoni *dataframe*-e kasutades andmeid salvestada mälus ja neid juba töödelda enne, kui need on kordagi kirjutatud Exceli faili. See võib oluliselt vähendada ajakulu suurte andmemahtude töötlemisel ja suurendada analüüsi efektiivsust. Seetõttu võib olla kasulik kaaluda VBA asemel mõne teise programmeerimiskeele kasutamist, kui töötlemist vajavad suured andmemahud.

Selle tarkvararakenduse suur puudus seisneb selles, et see ei suuda tekstist infot välja lugeda, vaid tugineb ainult olemasolevale andmebaasiinfole. Selleks saaks kasutada tekstikaevandamist, mille rakendamine on ka üks suurimaid edasiarendus võimalusi.

Näiteks on ettevõtte tegevusvaldkonna kindlaks tegemiseks oluline ettevõtte poolt registreeritud EMTAK kood ja majandusaasta aruandes jaotuses “Müügitulu jaotus tegevusalade lõikes” välja toodud põhitegevusala EMTAK koodi vastavuse kontrollimine. Seni kuni need EMTAK koodid on samad, ei ole rohkem küsimusi. Kui need koodid on aga erinevad oleks vaja neid kontrollida. Tekstilisel kujul olev tegevusaruanne majandusaasta aruandes on heaks kohaks nende kontrollimiseks. Kui tegevusaruanne on tehtud korralikult on seal ka välja toodud millega ettevõtte tegeleb. Kui tarkvara suudaks töödelda teksti ja näiteks otsida sealt tegevusaruandest märksõnu, mis käivad konkreetse raporti valdkonna kohta, siis oleks võimalik kontrollida kumb on õigem, kas ettevõtte registreeritud EMTAK või majandusaasta aruandes välja toodud EMTAK.

Tegevusaruandest oleks võimalik ka välja lugeda kas juhatuse liikmetele makstakse juhatuse liikme tasu. See võimaldaks korrigeerida palgaarvutust, sest hetke seisul deklareeritakse MTA-le kogu tööjõumaksude summa, kuid ainult töötajate arv ilma juhatuse liikmete arvuta [16]. Kui ettevõtte juhatuse liikmetele makstakse tasu, siis see moonutab ettevõtte keskmise palga arvutamist, sest juhatuse liikmed ei ole arvestatud töötajate hulka, kuid nende tasu on arvestatud tööjõumaksude hulka.

Lühidalt oleks tekstikaevandamine hea edasiarendus võimalus, et tarkvara suudaks töödelda teksti ja leida sealt olulist infot. Näiteks võiks tarkvara kontrollida ettevõtte registreeritud EMTAK-koodi vastavust majandusaasta aruandes välja toodud EMTAK-koodile ning tuvastada kas juhatuse liikmetele makstakse tasu.

## 5 Kokkuvõte

Äripäeva infopank pakub konkurentsiraporteid, mis annavad ülevaate konkreetse sektori ettevõtete finantsnäitajatest, turuosast ja -positsioonist. Praegu toimub konkurentsiraportite koostamine Excelis käsitsi, mis nõuab palju manuaalset tööd. Lõputöö eesmärk oli leida võimalusi selle protsessi automatiseerimiseks, et suurendada konkurentsiraportite kvaliteeti ja aidata klientidel strateegilisi otsuseid langetada. Tarkvaralahenduse loomine jagati neljaks etapiks: tutvuti olemasoleva lahendusega, uuriti ja planeeriti, pandi paika nõuded ning loodi prototüüp ja analüüsiti seda. Lähtudes konkreetse ettevõtte soovist, et kasutajad ei peaks rakenduse arendamiseks ja kasutamiseks alla laadima eraldi tarkvara sai valitud konkurentsiraportite andmeanalüüsi automatiseerimiseks VBA. Autor lõi rakenduse tegemiseks kaheksa erinevat moodulit, mis hõlpsustasid tööprotsessi rakenduse arendamisel ja aitavad tulevikus koodibaasi haldamisel koodis paremini orienteeruda. Rakendus jaotati viieks eraldiseisvaks osaks. Iga osa on rakendatav eraldiseisvatel Exceli lehtedel, kuhu on lisatud kasutusjuhend, mis aitab kasutajal programmi lihtsamini navigeerida. Autor otsustas valideerida tulemusi, võrreldes erinevate näitajate põhjal, kui kaua läheb aega andmefaili koostamisel käsitsi ja rakendust kasutades. Testide tulemuste põhjal oli uus lahendus vana lahendusega võrreldes ligi 12 korda kiirem ja efektiivsem. Kuid suurte andmemahutude töötlemine võib võtta palju aega, eriti kui kasutatakse VBA-d. Üheks suuremaks edasiarendusvõimaluseks on tekstikaevandamise kasutamine, mis võimaldaks lahendada kitsaskoha, et rakendus ei suuda infot tekstist välja lugeda.



## Kasutatud kirjandus

- [1] Riigi Infosüsteemi Amet, „Majandusaasta aruande esitamise tähtaeg,“ Eesti Firma, [Võrgumaterjal]. Available: <https://www.eestifirma.ee/et/majandusaasta-aruande-esitamise-tahtaeg/>. [Kasutatud 1 Mai 2023].
- [2] Eesti.ee, „Põhitegevusala valimine,“ Eesti.ee, [Võrgumaterjal]. Available: <https://www.eesti.ee/et/ettevotlus/ettevotte-loomine/pohitegevusala-valimine>. [Kasutatud 1 Mai 2023].
- [3] Riigi Infosüsteemi Amet, „EMTAK tegevusalad,“ Riigi Infosüsteemi Amet, [Võrgumaterjal]. Available: <https://www.rik.ee/et/e-ariregister/emtak-tegevusalad>. [Kasutatud 1 Mai 2023].
- [4] Microsoft, „What is SQL Server Management Studio (SSMS)?,“ Microsoft, [Võrgumaterjal]. Available: <https://learn.microsoft.com/en-us/sql/ssms/sql-server-management-studio-ssms?view=sql-server-ver16>. [Kasutatud 1 Mai 2023].
- [5] D. Mulhall, „Our Molecular Future: How Nanotechnology,“ New York, Prometheus Books, 2002, p. 392.
- [6] H. Zhou, „A Step-by-Step Approach for Understanding Machine Learning Methods,“ %1 *Learn Data Mining Through Excel*, Apress Berkeley, 2020.
- [7] J. Walkenbach, *Excel VBA Programming for Dummies 2nd Edition*, Indianapolis: Wiley Publishing, Inc., 2010.
- [8] T. Baier, E. Neuwirth ja M. De Meo, „Creating and Deploying an Application with (R) Excel and R,“ *R Journal*, kd. 3, nr 2, 2011.
- [9] T. Dysarz, „Application of Python Scripting Techniques for Control and Automation of HEC-RAS Simulations,“ *Advances in Hydraulics and Hydroinformatics*, kd. 10, nr 10, p. 1382, 2018.
- [10] G. Games, D. Witten, T. Hastie ja R. Tibshirani, *An Introduction to Statistical Learning with Application in R*, 2021.
- [11] S. M. Biju, „Comparative Analysis of Big Data Analytics Software in Assessing Sample Data,“ *Journal of International Technology and Information Management*, kd. 26, nr 2, pp. 1-15, 2017.
- [12] W. McKinney, *Python for Data Analysis*, Sebastopol: O'Reilly Media, Inc., 2013.
- [13] A. Nassereldine, P. Chen ja J. Xiong, „Excel Spreadsheet Analyzer,“ 2022.
- [14] A. H. Tan, „Text mining: The state of the art and the challenges,“ *Proceedings of the PAKDD Workshop on Knowledge Discovery from Advanced Databases*, kd. 8, pp. 65-70, 1999.
- [15] N. Bhanot, H. Singh, D. Sharma ja H. Jain, „Python vs. R: A Text Mining Approach for analyzing the Research Trends in Scopus Database,“ %1 *2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA)*, Tokyo, 2019.

[16] EMTA, „Statistika ja avaandmed: tasuta maksud üldinfo,“ EMTA, [Võrgumaterjal]. Available: <https://www.emta.ee/eraklient/amet-uudised-ja-kontakt/uudised-presseinfo-statistika/statistika-ja-avaandmed#tasatud-maksud-uldinfo>. [Kasutatud 14 Mai 2023].

## **Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks<sup>1</sup>**

Mina, Johanna Brigitte Verro

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Andmete analüüsi ja valimi kontrollimise automatiseerimine ettevõtte Äripäeva AS konkurentsiraportite näitel“, mille juhendaja on Inna Švartsman
  - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

17.05.2023

---

<sup>1</sup> Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtjaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.