

Group Introduction

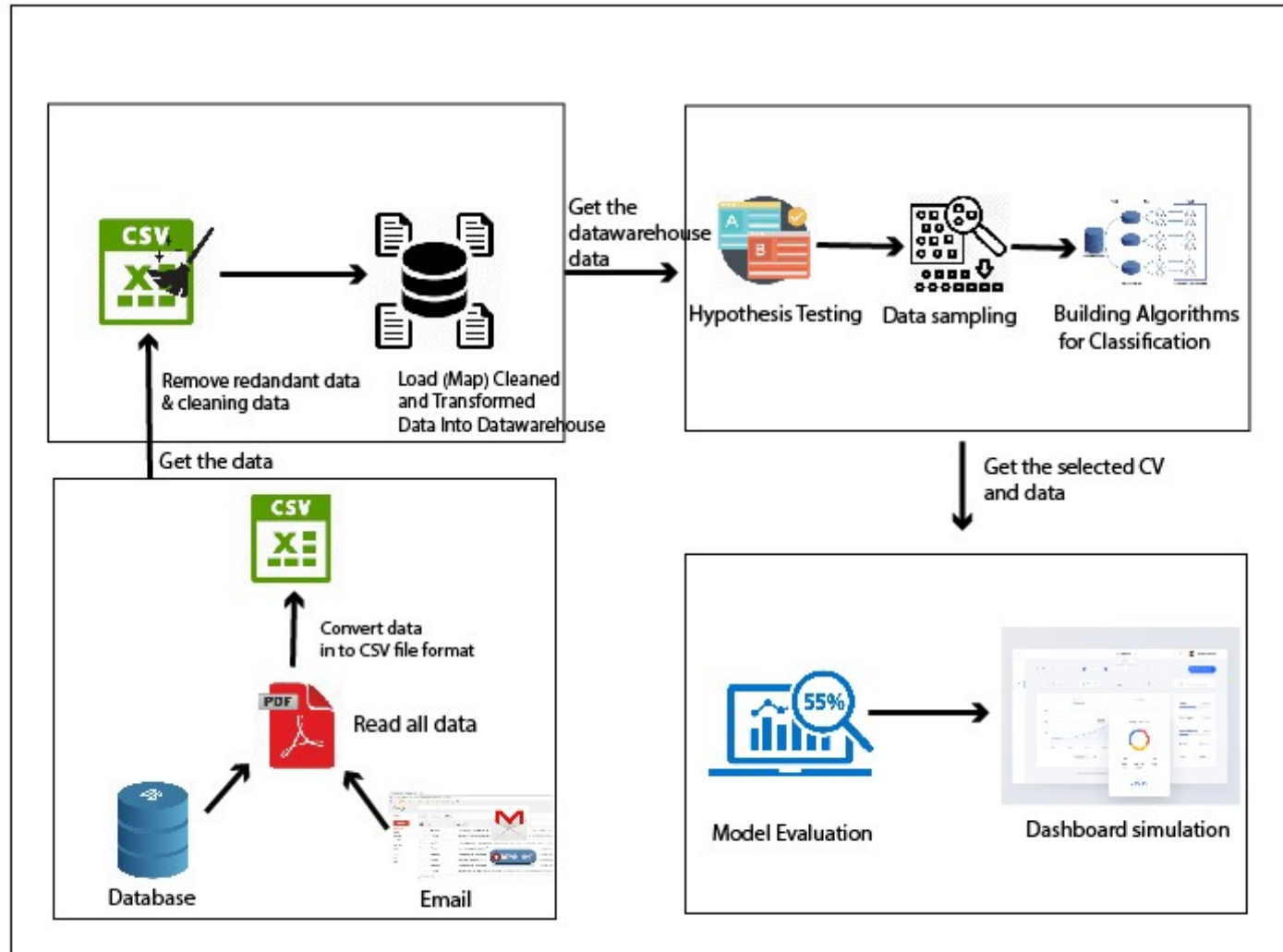
- Project Supervisor : Mr. Lakmal Rupasinghe
- Senarath S.P IT15099778
- Madhushika K.A IT15097798
- Yureshani H.B.D IT15040404
- Y.I. Kodithuwakku IT14115776

Research Gap

Features	Oracle BI	Birst	Jobscan	Smart Recruit	Smart Resume
BOT(Automated)					✓
ETL Tool					✓
Optimal Solution			✓		✓
Feasible Solution					✓
All the user does not have to interact with the system	✓	✓	✓	✓	✓
Visualization of data in an abstract way		✓	✓	✓	✓
Predictive Analytics	✓	✓	✓		✓

Objectives

System Overview



Research Problem



Individual Tasks

- | | |
|----------------------|---|
| 1. Senarath S.P | BOT Creation |
| 2. Madhushika K.A | Data warehousing using ETL Tool |
| 3. Yureshani H.B.D | Model Building and Prediction |
| 4. Y.I. Kodithuwakku | Model Evaluation and Dashboard Simulation |

1.BOT Creation

An Internet bot will be implemented to do the following tasks.

- Download the CVs in the email automatically
- To read the downloaded emails
- Classify the data in to relevant columns
- Save the classified data in CSV format

To implement the internet bot Neural Networks will be used over NLP (Natural Language Processing).

1.Download the CVs in the email automatically

- A novel email response algorithm

2.To read the downloaded emails

- The system that automatically read the emails,

Algorithms

- Multinomial Naive Bayes. This is a classic algorithm for text classification and natural language processing (NLP).
- Probability of class:

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w|c) = \frac{\text{count}(w,c) + 1}{\text{count}(c) + |V|}$$

3. Classify the data in to relevant columns

Apple CEO **Tim Cook** Introduces 2 New, Larger iPhones, Smart Watch At **Cupertino** **Flint Center** Event

Person

Organisation

Location

Figure 1: An example of NER application on an example text

4. Save the classified data in CSV format

- Save the Repository

2.Data Warehousing using ETL Tool

- Identify the Attributes with the use of Relevant data
- Extract the Relevant Data From CSV Files
- Cleaning Unorganized, Redundant Data (Data Cleansing)
- Transform Cleaned Data into Normalized Format
- Load (Map) Cleaned and Transformed Data Into Data warehouse

ETL Tool (Extract Transform Load)

- Process that extract data from external source then transform into normalize data format and load into data warehouse.



ETL Process

- Extract
 - Extracted the data for data transformation from source files.
 - Source file – CSV
(comma separated values)
 - Attributes -name
 - gender
 - age
 - email
 - Phone
 - qualifications
 - experience

A3	:	✖	✓	fx	2;"Mr";"Joyce";"Mills";"F";"jmills1@answers.com";28;"Software Engineer";"UptoDegree";9.8;0.0;0;"";8;2;7;7;3;6;2;1;4;4;4;1;1;3;5;4;3;1;5;0;0;0;"Strongly Agree";"No																▼
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	id	Title	first_name	last_name	emp_gender	email	age	Designation	educational_qualification	Working_experience_current	Working_experience_previous	is_pmp_certified	Background	Tech							
2	1	Mrs	Samuel	Armstrong	M	sarmstrong0@wunderground.com	41	Project Manager	UptoDegree	8.4;6.6;0	7;3;0;5;8;3;1;5;3;8;9;4;4;8;1;7;4;3;3;3	Disagree	No Opinion	Agree	No Opinion	Stro					
3	2	Mr	Joyce	Mills	F	jmills1@answers.com	28	Software Engineer	UptoDegree	9.8;0.0;0	8;2;7;7;3;6;2;1;4;4;1;1;3;5;4;3;1;5;0;0;0	Strongly Agree	No Opinion	Disagree	Agree	Agree	Agree	Dis			
4	3	Mr	Kathleen	Dixon	F	kdixon2@sbwire.com	39	Developer	Upto Advanced Diploma	2.6;10.4;0	3;7;5;4;7;4;3;4;1;4;2;2;3;4;2;8;2;6;5;6;3	Agree	No Opinion	Disagree	No Opinion	Agree	Stron				
5	4	Ms	Louise	Wright	F	lwright3@miibeian.gov.cn	45	Software Engineer	Upto MSc	14.3;0.7;0	3;3;4;5;3;1;3;4;3;5;4;4;3;1;2;6;5;4;0;3;0	Strongly Disagree	Strongly Disagree	Agree	No Opinion	Di					
6	5	Mr	Joan	Garcia	F	jgarcia4@cocolog-nifty.com	28	Trainee Developer	Upto Diploma	9.4;0.0;0	6;8;4;4;7;3;3;3;4;5;4;2;4;3;7;1;1;4;4;5	Strongly Agree	Agree	Strongly Agree	No Opinion	No Opi					
7	6	Mrs	Jennifer	Fuller	F	jfuller5@usnews.com	31	Team Lead	Upto MSc	8.3;3.7;0	4;8;3;3;2;1;4;4;8;7;3;0;5;1;0;2;5;5;6;1;2	No Opinion	Strongly Agree	Strongly Disagree	Strongly Disagree	Agree					
8	7	Ms	Beverly	King	F	bking6@ifeng.com	20	Developer	Upto Diploma	0.5;1.5;0	1;7;3;2;4;0;4;5;3;2;6;2;1;3;5;0;4;5;3;0;4;5	Agree	Disagree	Agree	No Opinion	Disagree	Disagree	Strongly Disag			
9	8	Mrs	Russell	Spencer	M	rspencer7@ed.gov	34	Software Architect	Upto MSc	13.1;0.0;0	8;4;5;1;7;5;3;2;4;2;2;3;4;5;2;1;6;7;0;2;3;3	Disagree	No Opinion	Disagree	Disagree	Disagree	No Opini				
10	9	Mr	Raymond	Dean	M	rdean8@sciencedaily.com	26	UI Engineer	Upto Diploma	7.7;0.0;0	6;2;3;8;4;8;1;0;2;8;8;0;1;1;5;6;8;2;5;2;5;6	Agree	Agree	Disagree	Disagree	No Opinion	Agree	Stron			
11	10	Ms	Lisa	Reynolds	F	lreynolds9@youku.com	34	Senior Software Engineer	Upto MSc	14.0;0.0;0	2;8;5;9;6;4;4;1;2;1;7;3;0;2;7;4;7;3;4;5;1	Strongly Agree	Agree	Disagree	Strongly Disagree	Str					
12	11	Mr	Mark	Stewart	M	mstewart@bloomberg.com	20	Data Entry Operators	Details Not Available	0.6;1.4;0	8;6;2;4;8;8;3;3;4;7;1;5;1;4;3;7;3;5;1;1;5;1	No Opinion	Agree	Strongly Disagree	No Opi						
13	12	Ms	Ruth	Payne	F	rpayneb@yahoo.com	34	Business Analyst	Upto Diploma	13.3;0.0;0	8;6;2;7;5;8;5;5;5;1;7;4;5;8;4;3;5;2;6;2;1	Disagree	Strongly Disagree	No Opinion	Agree	Disagree	Stro				
14	13	Ms	Paula	Bennett	F	pbennett@businessweek.com	20	Developer	Upto Diploma	1.0;0.0;0	9;4;4;8;7;0;1;1;2;9;4;1;0;6;2;5;3;2;2;1;4	Strongly Disagree	Disagree	Agree	Disagree	Disagree	S				
15	14	Mrs	Craig	Sanchez	M	csanchezd@bbc.co.uk	19	UI Engineer	Upto Diploma	0.8;0.0;0	3;7;1;2;3;2;0;1;3;2;8;5;3;3;6;4;8;1;5;4;0;4	No Opinion	Strongly Disagree	Strongly Disagree	Disagree	Agree					
16	15	Mr	Nicholas	Morgan	M	nmorgane@timesonline.co.uk	20	Developer	Upto Diploma	1.4;0.0;0	2;9;1;9;9;7;1;3;4;3;2;1;5;2;3;8;8;0;4;4;3;6	Agree	No Opinion	Agree	Strongly Disagree	Disagree					
17	16	Mr	Theresa	Stewart	F	tstewart@microsoft.com	19	Trainee Developer	Details Not Available	0.5;0.0;0	8;5;2;4;4;3;6;4;2;9;4;5;1;2;4;7;5;7;3;2;5;1	Agree	Disagree	Strongly Disagree	Agree	Disa					
18	17	Ms	Frances	Diaz	F	fdiazg@ed.gov	25	Web Designer	Upto Diploma	6.9;0.0;0	4;5;6;2;6;2;3;4;2;3;6;2;0;5;3;8;5;4;3;0;5	Agree	Disagree	Disagree	Strongly Agree	Agree	Strongly Disagree	A			
19	18	Mr	Ralph	Kelly	M	rkellyh@opensource.org	27	Associate Team Lead	UptoDegree	3.8;4.2;0	7;3;4;2;8;2;3;3;7;2;2;5;4;3;1;7;5;3;3;1;5	Agree	No Opinion	No Opinion	Agree	Disagree	Agree				
20	19	Mr	Alan	Perez	M	aperez@github.com	22	UI Engineer	Upto Diploma	4.0;0.0;0	9;5;0;7;8;4;0;1;4;4;9;7;3;2;8;8;5;1;4;2;2;4	Agree	Agree	Strongly Disagree	Strongly Disagree	Agree	Strongly Di				
21	20	Ms	Ruby	White	F	rwhitej@bizjournals.com	26	Developer	Upto Advanced Diploma	2.6;4.4;0	6;6;4;1;6;6;0;5;1;6;5;7;3;4;2;1;4;2;4;1;4	Strongly Disagree	Strongly Disagree	Strongly Agree	Disag						
22	21	Mrs	Matthew	Perry	M	mperryk@google.fr	19	Data Entry Operators	Details Not Available	0.6;0.0;0	9;9;4;9;5;7;7;5;2;8;4;7;0;3;4;1;6;5;0;2;4;4	Strongly Disagree	Agree	No Opinion	Agree	Disag					
23	22	Ms	Phyllis	Watkins	F	pwatkins@noaa.gov	18	Trainee Developer	Upto Diploma	0.3;0.0;0	9;8;2;8;7;8;4;1;3;4;5;6;3;5;2;8;8;7;2;5;6;1	Disagree	Disagree	Disagree	No Opinion	Disagree	Disagr				
employee_data																					
READY																					

ETL Process

- Transform
 - Apply transformation for extracted data
 - Cleaning – e.g., mapping NULL to 0 or Male to M and Female to F
 - Filtering – select certain column to load
 - Splitting column to multiple columns and vice versa e.g. first name, last name
 - Applying data validations
 - Remove redundant data

ETL Process

- Load
 - Map the extract, transformed data into data warehouse for further processing
 - Load can be performed
 - Overwrite existing information
 - Updating extracted data frequently, daily, weekly etc.

eid	Full_name	Phone_1	Phone_2	Email	Gender	Qualification_1	Qualification_2	...
-----	-----------	---------	---------	-------	--------	-----------------	-----------------	-----

Single table

Data warehouse

- Place that store both historical and current data in a single place use for further data analysis and decision making
- Core components of the business intelligence

MySQL Data warehouse

- Data warehouse solution
- Open source
- Support for Data marts