VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Belagavi-590018

SRI JAYACHAMARAJENDRA COLLEGE OF ENGINEERING

(An Autonomous Institution)

Mysuru-570006

A Project Report on

Modular Survey System with Data Analytics

Submitted in partial fulfillment of the award of the degree of

Bachelor of Engineering (B.E.) in

Electronics and Communication Engineering

Submitted by

Ankita P	4JC13EC015
Ashish Gupta	4JC13EC019
Sandesh Jagadish	4JC13EC042
Sunil Kumar Y	4JC12EC420

Under the guidance of

Dr. M.N Shanmukha Swamy

Professor HOD, Dept of E & C,

SJCE,Mysuru

April 2017

JSS MAHAVIDYAPEETHA

Sri Jayachamarajendra College of Engineering

JSS S&TU Campus, Manasagangothri, Mysuru - 570006

Department of Electronics and Communication Engineering

Certificate

Certified that the project work entitled **'Modular Survey System with Data Analytics'** is a bonafide work carried out by **Ankita P, Ashish Gupta, Sandesh Jagadish** and **Sunil Kumar Y** in partial fulfillment for the award of the degree of **Bachelor of Engineering** in **Electronics and Communication** of **Visvesvaraya Technological University, Belagavi**, during the year 2017. It is certified that all corrections/suggestions indicated for the internal assessment have been incorporated in the final report deposited in the department library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the degree of Bachelor of Engineering.

Signature of HOD & Guide	**Signature of Principal**
Dr. M N Shanmukha Swamy	Dr. T N Nagabhushan

External Viva:

Name of the Examiners	**Signature with Date**
1.	
2.	

DECLARATION

We, Ankita P, Ashish Gupta, Sandesh Jagadish and Sunil Kumar Y, the students of final semester B.E., Electronics and Communication Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, hereby declare that the project work entitled Modular Survey System with Data Analytics has been carried out by us under the guidance of Dr.M.N Shanmukha Swamy , Professor and Head , Department of E & C, SJCE for the partial fulfillment of the curriculum prescribed for the award of the degree of Bachelor of Engineering in Electronics and Communication Engineering by the Visvesvaraya Technological University, Belagavi, during the 2016-2017 term.
We also declare that, to the best of our knowledge and belief, the matter embodied in this project report has not been submitted previously for the award of any degree to any other institution or university.

Place: Mysuru

Date:

Ankita P (4JC13EC015)

Ashish Gupta (4JC13EC019)

Sandesh Jagadish (4JC13EC042)

Sunil Kumar Y(4JC12EC420)

ACKNOWLEDGEMENT

The success of any endeavor depends a lot on the goals set at the onset as well as the constant guidance and motivation received throughout. We take this opportunity to express our deepest gratitude and appreciation towards all those who have helped us directly or indirectly with the successful completion of this project.

First, we would like to thank Dr. T N Nagabhushan, Principal, SJCE. We would also like to thank Dr. M. N. Shanmukha Swamy, HOD, Department of Electronics and Communication, SJCE, Mysuru for providing the opportunity to work on this exciting project. We would like to thank the project coordinator, Prof. V. Nattarasu and the panel members, Dr. M. N. Shanmukha Swamy, Mrs Anitha S Prasad and Dr. N. M. Renukappa for helping us with their invaluable feedback and insight.

We would like to express our deep sense of respect and gratitude towards our advisor and guide Dr. M. N. Shanmukha Swamy, Professor and Head, Department of Electronics and Communication Engineering, who has been the guiding force behind this work. We are greatly indebted to his for his constant encouragement, valuable advice and for propelling us further in every aspect of our academic life. His presence and optimism have provided an invaluable influence on our career and outlook for the future.

Finally, we would like to extend our gratitude to all those who have indirectly helped us complete this project successfully.

ABSTRACT

In this era of ever-increasing data collection and voluminous data interpretation, it is inevitable to generate self-sufficient and modern systems. Satisfying the need for data analysis is essential in today's world. We aim to achieve this by adopting a modular approach. Our proposed device is designed to adapt and carry out any variation of census, voting or simple data acquisition, with the help of user friendly code snippets that integrate the modular snap-on. Our project addresses this issue of massive data analysis by developing a custom survey system for modes of voting and polling. The system developed is made very convenient to operate and to generate the analysis. Our project involves the use of latest technologies in use as of now. The system is adaptable and customizable. The problems faced in this industry is addressed in our project by developing a system which can be a replacement of the current EVMs in the long term. This system offers adaptability to any voting/data acquisition process, high level of security, basic and efficient user interface, data analysis and prediction and portability. We intend to integrate all these features and develop a unique and modular system.

Contents

List of Figures

Chapter 1

INTRODUCTION

If there is something that has been traversing at a rapid rate on the exponential curve, it has to be the acquisition of data in the everyday functioning of the world. Data is an essential component in today's technological world being integrated in all walks of life. Different fields ranging from art and science to politics, sports and social media, all involve the acquisition and collection of data. There is a great need to organize and classify this data in the way that a layman can understand. Data analysis in this age is required to influence outcomes for the future and to interpret different models of current technologies in place.

1.1 Problem Definition

With the advancements in science and technology and with the ever increasing ease of mass production, one prefers to manufacture a purpose-centric device to perform a specific task. Most often these tasks are not perennial, hence the actual productive period of such devices amounts up to a small fraction, thus questioning convenience over efficiency. A meaningful approach is to emphasis on the robustness of the device to adapt to multiple tasks without compromising the convenience offered by the purpose-centric devices.

The Electronic voting machines (EVMs) currently in use in place of the physical ballot papers, since 1999. It consists of 2 systems, a Balloting unit - a unit used by the voters to enter their choice, and a Control unit - to record and store the votes. The current EVMs do not transmit the data to a centralized server and each device display the result individually. They are hard-coded for security concerns. This means that the device cannot be used / reconfigured for any other purpose. This drastically increases the idle time of the device. The present day value of a new EVM is around 40000 per device, which is not economical, given that its used only a handful times per year and they are relatively bulky. A single unit can record and store only 3480 votes.

1.2 Literature survey

Storing critical data secure and accessible from several locations has become a global goal, may it be personal data, organizational or data from applications. An innovative approach to this issue lead to the emergence of Cloud Computing, that allow users to store and access data in the Cloud - remotely accessible storage servers.

Joo Tiago and Medeiros Paulo have presented a good insight about the task at hand in their paper - Efficient Storage of Data in Cloud Computing. They explain how an efficient and secure management of the Cloud services, can improve the quality of services being offered, optimize the resources and infrastructures required and how it can be scaled easily compared to any other mode of secure storage. After stating the beneficial grounds of Cloud Computing, they present a solution to avoid and eliminate data redundancies between virtual machines on shared data.

Our project emphasizes on analytics performed on data collected through the customizable survey system. The paper 'Solving the Internet of Things Last-Mile Problem' by 'Vaidy Krishnan, Tableau Software' depicts the growth on haphazard data on a large scale. The connectivity of devices is not a problem anymore, the current issue is understanding of the data collected by all devices.

With internet and rapid communication growth worldwide, the data is exponentially increasing and it is inevitable to develop smart systems to analyse the ever growing data. The paper suggests methodologies which are relevant to the process of analytics used in our project. Fast iteration and deeper interconnectivity of data play an important role in determining the route necessary for approaching of the collected data. Data blending, sharing and collaboration enhances the executive speed at which analysis takes place. Our project implements such techniques for improving the functionality of the customizable survey system.

1.3 Organization of the Project Report

The report has been organized as follows:

CHAPTER 2

Provides information on the proposed concept, implementation methodology and the applications of the designed system.

CHAPTER 3

Lists the hardware and software requirements for developing the survey prototype and model.

CHAPTER 4

Explains the design and development of the box-build prototype.

CHAPTER 5

Gives a detailed explanation on the integration and interfacing of software and hardware along with the functions used while coding.

CHAPTER 6

Provides basic information on the topic of data analytics, machine learning and the mathematics involved in our project.

CHAPTER 7

Demonstrates the output graphs obtained and the execution of MATLAB and python scripts.

CHAPTER 8

Gives a brief summary of our project and lists the further scope and future use of the proposed concept.

Chapter 2

SYSTEM OVERVIEW

2.1 Proposed Concept

Our project undertakes this issue regarding the data conundrum to come out with a better solution to data management. The idea is to develop a customisable survey system which is user-friendly, portable and can be used for wide-ranging applications. Our project takes this up a notch by adding the current trends in technology like security and analytics with effective implementation.

The deployment of the device involves a predefined set of steps, which remains the same for any task its being programmed for. A flow diagram illustrating the steps involved in the process is shown below.

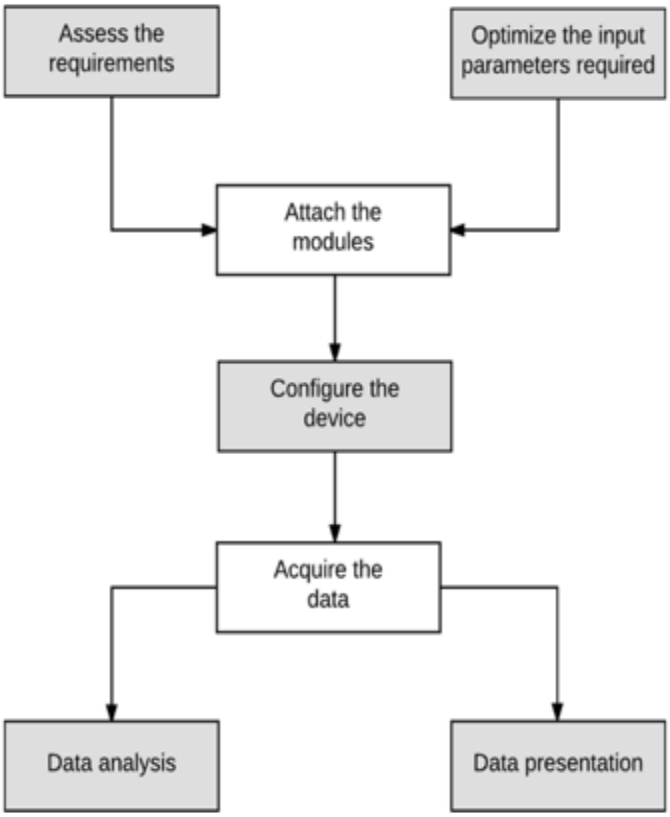

Figure 2.1: Flow diagram showing employment of the proposed device

2.2 Implementation of the Survey Mechanism

1. **Voting:**

   Voting is the mode where one user selects an option among several options, usually done in an election process. Voting involves the analysis of best option and the worst option selected by the users of the survey system along with the result distribution for each option.

   The device currently provides provision for 8 candidates to demonstrate the usage of the product for vote recording during an election. The number of candidates supported can easily be extended by a using an external keypad along with a codec pair to communicate with the device. Upon validated entry into the program, in each run the programs checks for an input registered and identifies a key pressed. Corresponding functions are called to update the databases and finally record the changes onto the SD card at the end of the cycle.

2. **Polling:**

   Polling is the other mode in our survey system where each user answers multiple questions and the data is developed for interpreting an opinion or an idea.

   The current model offers facility simple Yes/No replies. The result is stored in a 2D matrix q_status[n][4], here n refers to the number of questions one intends to configure the device with. Push buttons are used to record a response. Digital pins are configured as Pull up by default. Clicking of the push button grounds the digital pin corresponding to an option, which is scanned by the poll() function.In later stages, the acquired data will be stored in an on board SD card, which shall be used to transfer the data onto a system for further processing. Gathered data is analyzed using suitable data mining algorithm on MATLAB platform. The modules developed later represents the analyzed data through pie charts and bar graphs, and using the current data set, predict the outcome of the key question based on the parameters entered.

   The survey system provides easy implementation of both voting and polling modes. The procedure of execution and the operation of the system remains the same. Hence, a variety of surveys can be conducted by anyone and the data can be easily interpreted.

2.3 Applications

The applications of this device in its respective field is potentially limitless. It integrates the functionalities of all its counterparts into a single, portable device. Some of the everyday applications are mentioned below:

1. **Enumeration and census** - provide an insight into the current status of related parameters and serves as the basis of constructing, planning, forecasts. It is also an essential part for a democratic process as it enables the citizens to examine the decisions made by the government and the local authorities and decide their effectiveness. But every enumeration process comes with an inherent time gap between its iterations, during which the corresponding devices remain idle. With the adoption of the proposed device, a single device can be configured to be used with any enumeration platforms and hence mitigating the idle time.

2. **Polling** - The most strategically precise move before taking a decision affecting the masses, is to take a poll of the opinions. The contents of such a poll are usually diverse, all of which can be configured on our device, leading a step closer towards an efficient system. In this opinionated world, the system offers everyone a chance to voice their ideas and opinions. This is the fundamental concept of the democracy that we live in.

3. **Customized Data Entry** - Camera module, fingerprint scanner and the included hex keypad module can be configured to form a basic data acquiring device, that can easily be appreciated for its convenient multipurpose in Aadhaar office, Passport centres, RTOs etc. Any layman can understand the operation of the system and this would enable easy customization of the system providing multivariate surveys for everyone to interpret and understand.

4. **Prediction** - Another important application to the proposed concept is the prediction model developed. Prediction is an essential part in various fields of study which help preparing for future consequences. The data collected from the survey is subjected to machine learning algorithms to predict using a probabilistic approach as to what the output can be provided the same features.

Chapter 3

HARDWARE AND SOFTWARE REQUIREMENTS

3.1 Hardware Requirements

3.1.1 Arduino Uno R3

The Arduino Uno is a microcontroller board based on the ATmega328 (datasheet). It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header, and a reset button. It contains everything needed to support the micro-controller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started. The Uno differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it features the Atmega16U2 (Atmega8U2 up to version R2) programmed as a USB-to-serial converter.

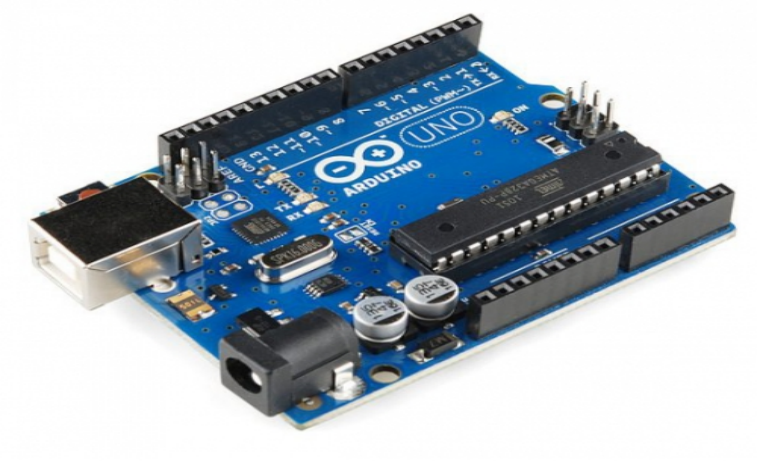

Figure 3.1: Arduino Uno Board

3.1.2 Arduino Mega 2560

ATmega2560 based microcontroller, the Mega 2560 has 54 digital input/output pins, 16 analog pins, a 16 MHz crystal oscillator, a USB connection, a power jack, ICSP header and a reset.

Mega is compatible with most shields made for the Uno and former boards. Working with both the boards requires 2 elementary functions, which are:

Void Setup() - To initialize the variables and initial conditions. and contains the code to be run once, at the beginning of the program execution.

Void loop() - Contains the part of the code to be run continuously in a loop.

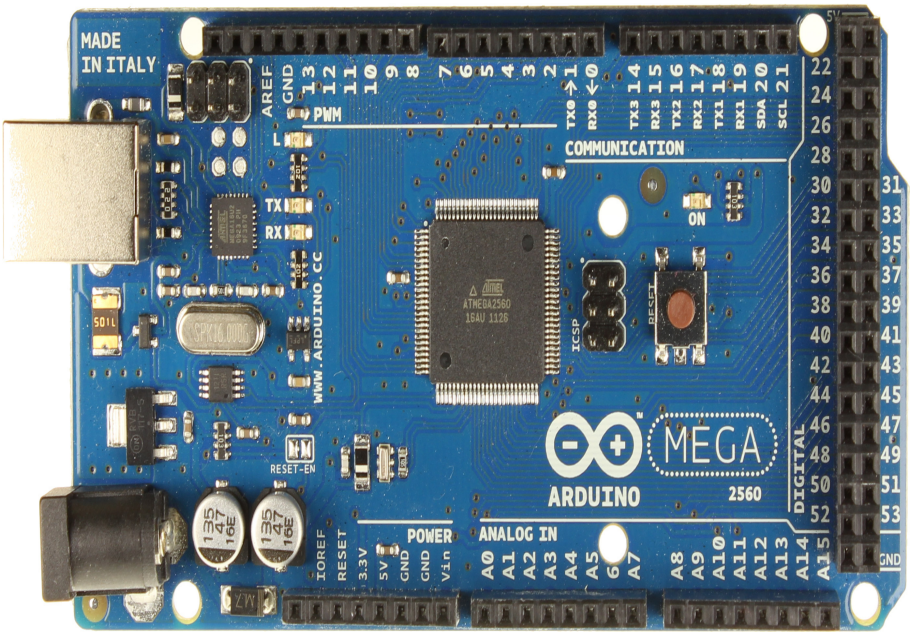

Figure 3.2: Arduino Mega Board

3.1.3 Display

LCD (Liquid Crystal Display) screen is an electronic display module and find a wide range of applications. A 16x2 LCD display is very basic module and is very commonly used in various devices and circuits. These modules are preferred over seven segments and other multi segment LEDs. The reasons being: LCDs are economical; easily programmable; have no limitation of displaying special even custom characters (unlike in seven segments), animations and so on.

A 16x2 LCD means it can display 16 characters per line and there are 2 such lines. In this LCD each character is displayed in 5x7 pixel matrix. This LCD has two registers, namely, Command and Data. The command register stores the command instructions given to the LCD. A command is an instruction given to LCD to do a predefined task like initializing it, clearing its screen, setting the cursor position, controlling display etc. The data register stores the data to be displayed on the LCD. The data is the ASCII value of the character to be displayed on the LCD.

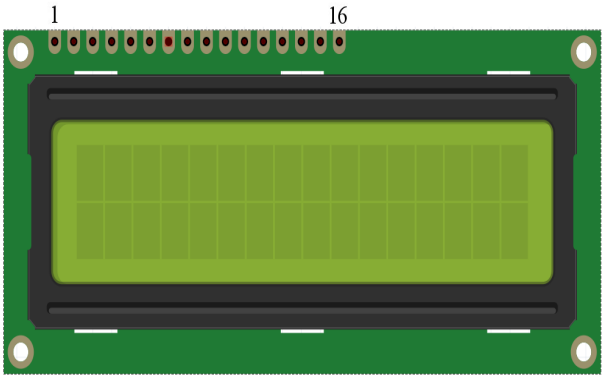

Figure 3.3: Liquid crystal display (16*2) used to display the survey questions

3.1.4 Buttons

A push-button or simply button is a simple switch mechanism for controlling some aspect of a machine or a process. Buttons are made out of hard material, usually plastic or metal. The surface is flat or shaped to accommodate the human finger or hand, so as to be easily depressed or pushed. Buttons are most often biased switches, though even many unbiased buttons (due to their physical nature) require a spring to return to their un-pushed state.

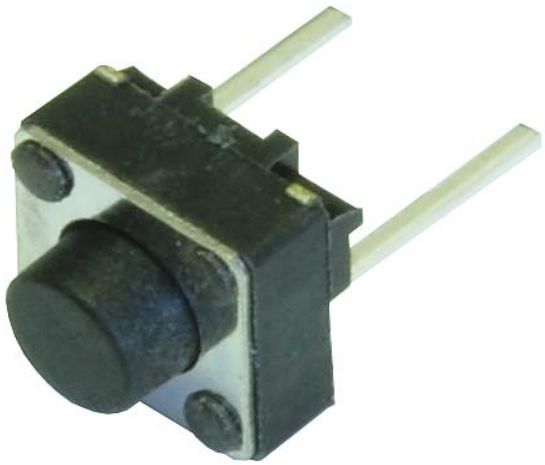

Figure 3.4: Push button used to input the user feedback

3.1.5 RFID modules

Radio-frequency identification (RFID) uses electromagnetic fields to automatically identify and track tags attached to objects. The tags contain electronically stored information. Passive tags collect energy from a nearby RFID readers interrogating radio waves. Active tags have a local power source such as a battery and may operate at hundreds of meters from the RFID reader.Unlike a bar code, the tag need not be within the line of sight of the reader, so it may be embedded in the tracked object The MFRC522 is a highly integrated reader/writer IC for contact-less communication at 13.56 MHz. The MFRC522s internal transmitter is able to drive a reader/writer antenna designed.Module uses voltage of 3.3V, through the SPI interface simple few lines can be directly connected to the user any CPU board communication module can guarantee stable and reliable work, reader distance.

Figure 3.5: RFID module used for providing admin rights and security

3.1.6 SD Card Adapter

The Arduino SD Card Shield is a simple solution for transferring data to and from a standard SD card. The pinout is directly compatible with Arduino, but can also be used with other microcontrollers. The other features include easy placement of SD card, sits directly on an Arduino and contains a switch to select the flash card slot.

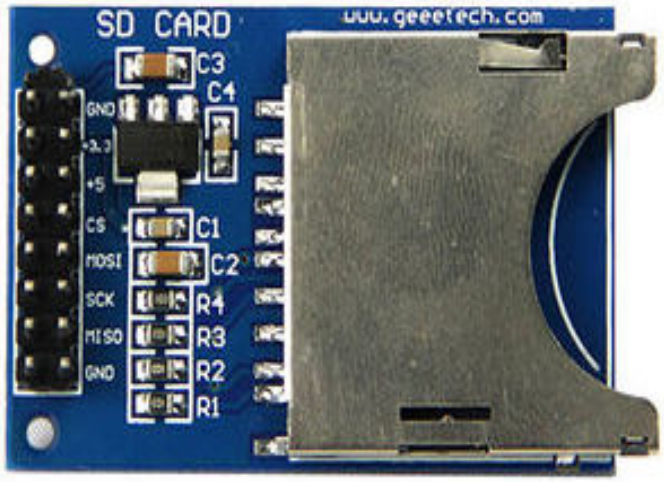

Figure 3.6: SD Card module used for storing survey data

The communication between the microcontroller and the SD card uses SPI, which takes place on digital pins 11, 12, and 13 (on most Arduino boards) or 50, 51, and 52 (Arduino Mega). Additionally, another pin must be used to select the SD card. This can be the hardware SS pin - pin 10 (on most Arduino boards) or pin 53 (on the Mega) - or another pin specified in the call to SD.begin().

3.2 Software Requirements

3.2.1 MATLAB

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment and fourth-generation programming language. A proprietary programming language developed by MathWorks, MATLAB allows matrix manipulations, plotting ounctions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, Fortran and Python. Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing abilities. An additional package, Simulink, adds graphical multi-domain simulation and model-based design for dynamic and embedded systems.

The MATLAB application is built around the MATLAB scripting language. Common usage of the MATLAB application involves using the Command Window as an interactive mathematical shell or executing text les containing MATLAB code. MATLAB supports developing applications with graphical user interface (GUI) features. MATLAB includes GUIDE (GUI development environment) for graphically designing GUIs. MATLAB can call functions and subroutines written in the programming languages C or Fortran.

Libraries written in Perl, Java, ActiveX or .NET can be directly called from MATLAB, and many MATLAB libraries (for example XML or SQL support) are implemented as wrappers around Java or ActiveX libraries.

Figure 3.7: MathWorks MATLAB used for analysis

3.2.2 Scikit-Learn

Scikit-learn is a free software machine learning library for the Python programming language.It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. SciKits (short for SciPy Toolkits), are add-on packages for SciPy, hosted and developed separately and independently from the main SciPy distribution. All SciKits are licensed under OSI-approved licenses.Packages are packaged as toolkits (instead of in the main, monolithic SciPy distribution) when:

1. The package is deemed too specialized to live in SciPy itself.
2. The package has a GPL (or similar) license which is incompatible with SciPys BSD license .
3. The package is meant to be included in SciPy, but development is still in progress. The advantage that consistent naming brings is that the package becomes easier to discover.

The technical step to make this happen is simple: register the project in the Python package index with a name chosen to start with scikit. Examples of existing package: scikit-learn (machine learning), scikits.sparse (additional sparse matrix routines).

Whilst the recommended license for Scikits projects is the (new) BSD license, scikits packages are free to choose their own open source license. The license should be officially OSI approved. We, the scipy-developers, will allow packages to contain code with licenses that, in our judgment, comply with the Open Source Definition but have not gone through the approval process. This is to allow us to adopt old code with permissive licenses. The package itself, though, should use a well-known OSI-approved license.

Figure 3.8: Scikit-learn used for python script prediction

3.2.3 Arduino IDE

Arduino is an open source, computer hardware and software company, project, and user community that designs and manufactures microcontroller kits for building digital devices and interactive objects that can sense and control objects in the physical world. The project's products are distributed as open-source hardware and software, which are licensed under the GNU Lesser General Public License (LGPL) or the GNU General Public License (GPL), permitting the manufacture of Arduino boards and software distribution by anyone. Arduino boards are available commercially in preassembled form, or as do-it-yourself kits.

Arduino board designs use a variety of microprocessors and controllers. The boards are equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits. The boards feature serial communications interfaces, including Universal Serial Bus (USB) on some models, which are also used for loading programs from personal computers. The microcontrollers are typically programmed using a dialect of features from the programming languages C and C++.

The Arduino project provides the Arduino integrated development environment (IDE), which is a cross-platform application written in the programming language Java. It originated from the IDE for the languages Processing and Wiring. It includes a code editor with features such as text cutting and pasting, searching and replacing text, automatic indenting, brace matching, and syntax highlighting, and provides simple one-click mechanisms to compile and upload programs to an Arduino board. It also contains a message area, a text console, a toolbar with buttons for common functions and a hierarchy of operation menus.The Arduino IDE supports the languages C and C++ using special rules of code structuring. The Arduino IDE supplies a software library from the Wiring project, which provides many common input and output procedures.

Key features:

1. Inexpensive and flexible hardware platform
2. Simple programming environment
3. Cross-platform-The Arduino Software (IDE) runs on Windows, Mac OSX, and Linux operating systems.
4. Open source and extensible software

Figure 3.9: Arduino IDE platform used for hardware interfacing

Chapter 4

PROTOTYPE DESIGN AND DEVELOPMENT

The components selected for the development of the project were required to be assembled in the most optimized way possible so as to produce an elegant user-friendly system. A box-build approach was considered for portability and efficient component placement.

The box was designed by using 3mm acrylic as the material. Holes were drilled through the box for the placement of components like the LCD, buttons, LEDs and the RFID interface.The dimensions were selected so as to provide clearance for wiring and other connections. Arduino Uno was connected to the RFID module. Arduino Mega was connected to the buttons and the LCD display. These boards were placed at the center of the base of the box. Batteries were added at the ends of the box. RFID module and LCD were suspended on the top face of the box for better interfacing and display respectively. The voting buttons were placed on the left side of the top face of the box providing upto 8 options. The YES/NO buttons were placed on the right side of the top face. The arrangement of the components of the top face is shown in the figure 7.2.

The dimensions of the box-build for the optimized arrangement of components and portability is shown in figure 7.1.

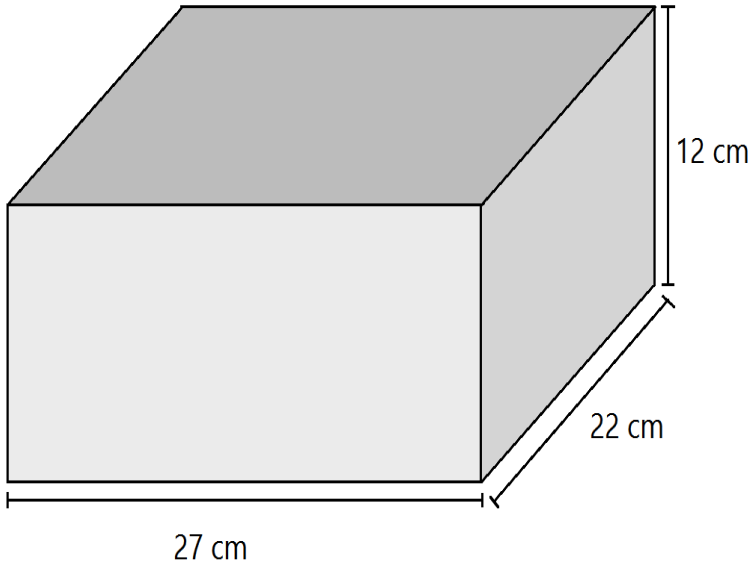

Figure 4.1: Dimensions used for design of the box-build

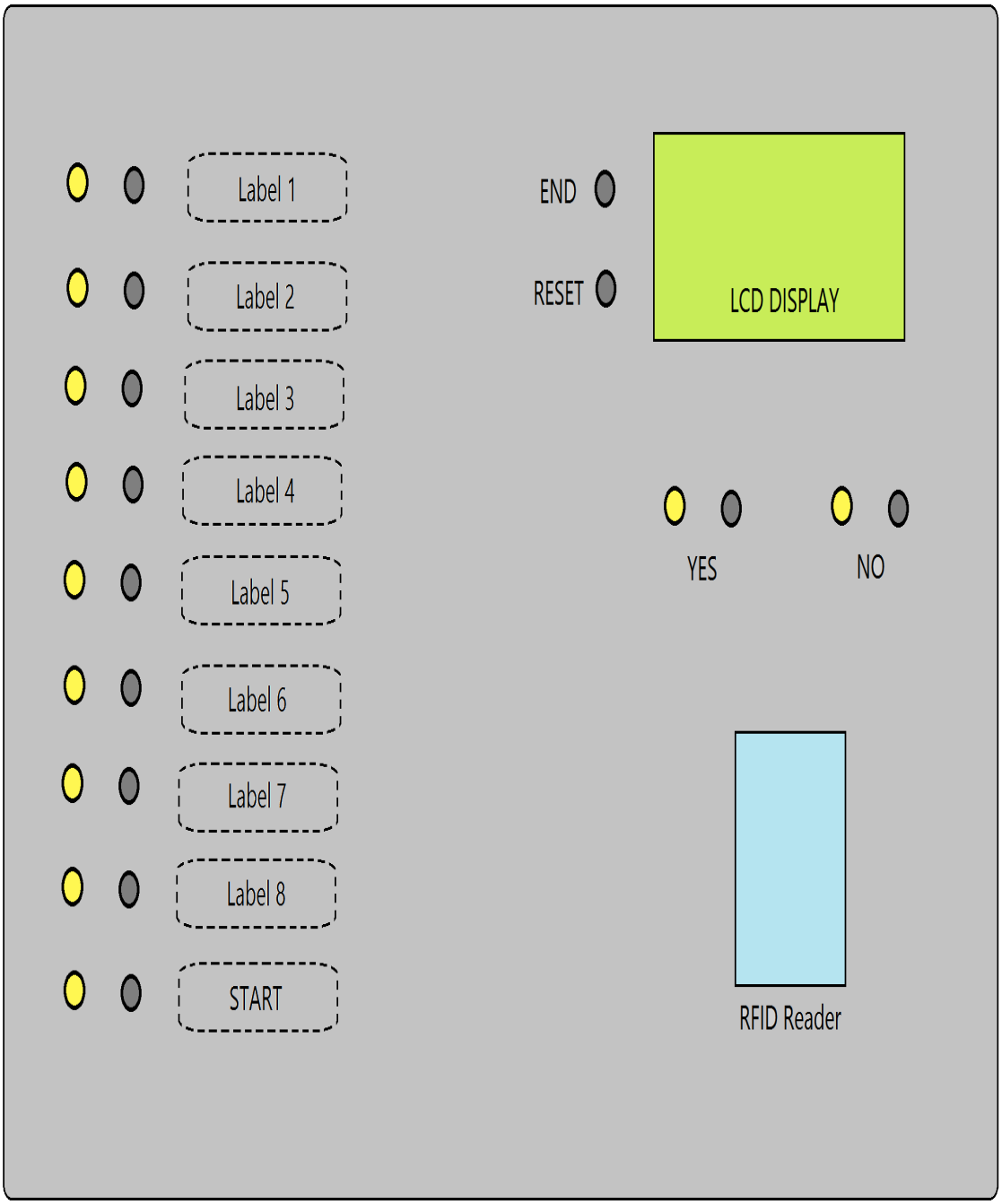

Figure 4.2: The top view of the designed survey system

Chapter 5

INTERFACING OPERATIONS

Since our proposed idea involves a fair share of both hardware and software operations, the interface to connect, read and then transmit the data to the softwares requires stability and efficacy. The functions used in the code and the methodology implemented to integrate hardware and software is elaborated in the below sections.

5.1 Hardware Section

5.1.1 User Defined Methods

1. void setup - To initialize the variables and contains the code to be run once, at the beginning of the program execution.
2. Void loop() - Contains the part of the code to be run continuously in a loop.
3. Void buzzer() - Triggers the buzzer to indicate the acceptance on an input.
4. Void results() - Displays the votes recorded for every question, and the options within a question, or the result of each candidate in case of voting scenario.
5. Void poll() - Scans through the switch network to detect any vote registered. Initiates the corresponding changes to be made in the database.
6. Void authenticate() - Authenticates the entry during start up, to end the survey and resetting the memory.
7. Void write_sd() - Interface with the SD module and record the required changes on to the SD card in a .CSV file.

5.1.2 Pin Details

Arduino end (Pin #)	Other end	Arduino end (Pin #)	Other end
2	LCD pin 14	44	Button Yes
3	LCD pin 13	45	Button No
4	LCD pin 12	46	LED Yes
5	LCD pin 11	47	LED No
6	LCD pin 6	48	Button reset
7	LCD pin 4	49	Button end
26	Button a	50	MISO
27	Button b	51	MOSI
28	Button c	52	SCK
29	Button d	53	CS
30	Button e		
31	Button f	UNO:	RFID:
32	Button g	10	SDA
33	Button h	13	SCK
42	Button start	11	MOSI
		12	MISO
14	Button authentication	9	RST
34	LED a	39	LED f
35	LED b	40	LED g
36	LED c	41	LED h
37	LED d	43	LED start
38	LED e		

Figure 5.1: Pin Description

5.1.3 LCD Interface

The display of choice is a 16x2 LCD screen, which offers an economical and easily programmable display solution. LCD has two registers- Command and Data registers. Command register stores the instructions given to a LCD and the Data register stores the data to be displayed, which passed in terms of the ASCII values. Power supply is given from the board to power the screen and the back light. Data pins D0 to D3 handles the data transfer to the screen and RS/RW specifies the reading or writing mode.

Functions used to interface LCD:

1. LiquidCrystal lcd(7,6, 5, 4, 3, 2) - declares the arduino pins to be used to control data and reset of the LCD.
2. lcd.clear() - Clears the content on the LCD.
3. lcd.setCursor(x,y) - Set the cursor coordinates to be accessed on the LCD screen.
4. lcd.print() - Print the passed content on the LCD screen, starting from the cursor position.
5. lcd.seek(int) - Set the cursor location to the passed location.
6. lcd.peek(int) - Retrieve the current cursor position.

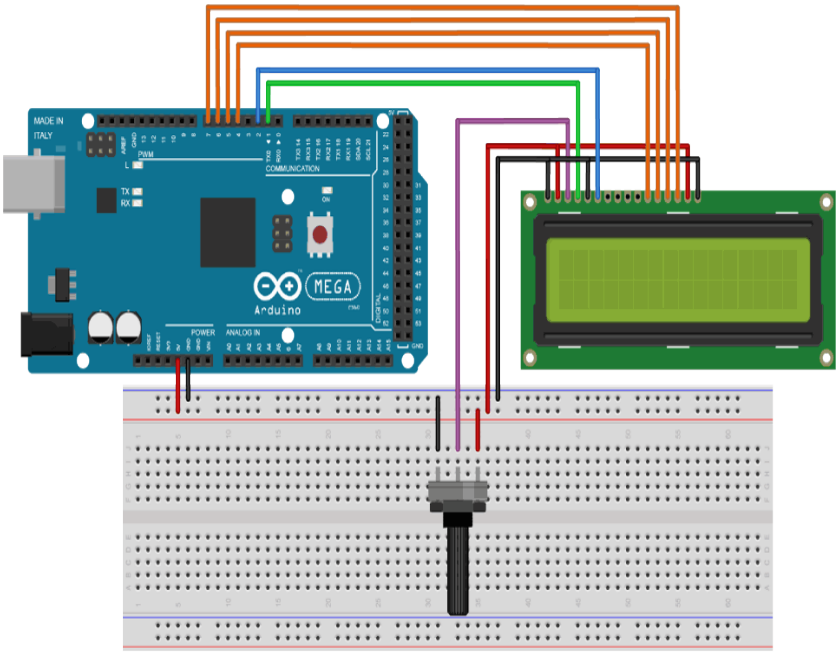

Figure 5.2: LCD interface with Arduino Mega controller

5.1.4 SD Card Module

The Arduino SD Card Shield is a simple solution for transferring data to and from a standard SD card. The pinout is directly compatible with Arduino, but can also be used with other micro-controllers. It allows you to add mass storage and data logging to your project.

The communication between the microcontroller and the SD card uses SPI, which takes place on digital pins 11, 12, and 13 (on most Arduino boards) or 50, 51, and 52 (Arduino Mega). Additionally, another pin must be used to select the SD card. This can be the hardware SS pin - pin 10 (on most Arduino boards) or pin 53 (on the Mega) - or another pin specified in the call to SD.begin().

Functions used to interface SD card:

1. SD.begin() - To setup a connection with the connected SD memory card.
2. SD.exists(filename.txt) - Returns true or false based on the existence of the given text file in memory.
3. SD.open() - Opens a file on the SD card. If the file is opened for writing, it will be created if it doesn't already exist (but the directory containing it must already exist).
4. SD.remove() - Removes the mentioned text file from the SD card.
5. SD.close() - Close the file, and ensure that any data written to it is physically saved to the SD card.
6. SD.println() - Print data, followed by a carriage return and newline, to the File, which must have been opened for writing. Prints numbers as a sequence of digits, each an ASCII character.
7. SD.seek() - Seek to a new position in the file, which must be between 0 and the size of the file (inclusive).
8. SD.position() - Get the current position within the file (i.e. the location to which the next byte will be read from or written to).
9. SD.read() - Read the next byte from the file linked to and stores in the buffer.

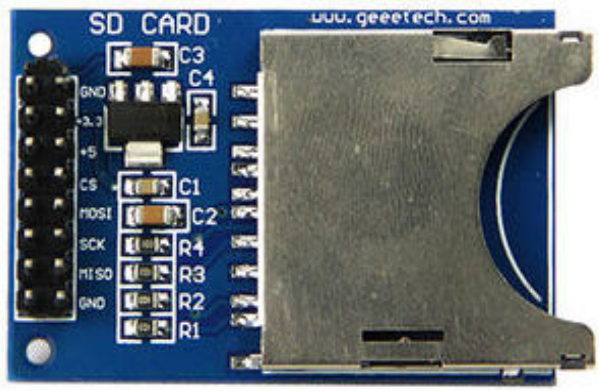

Figure 5.3: SD card module used for storing data in SD card

5.1.5 Serial Interface

Functions used to access serial port:

1. Serial.begin(9600) - Set up connection with the serial port with a baud rate of 9600bps. Included in the setup() function of the arduino code.
2. Serial.available() - Checks whether there is an incoming bit stream through the serial port.
3. Serial.read() - Reads the incoming data from the serial port byte wise.
4. Serial.write() - Send byte wise data into to the serial port.
5. Serial.print() - Write/display content on the serial window for debugging purpose.

Figure 5.4: Screenshot of the serial interface in the Arduino IDE

5.1.6 RFID Reader RC522

RC522-AN module adopts Philips MFRC522 original reader circuit chip design, easy to use, low cost, suitable for equipment development, development of advanced applications such reader users, the need for RF card terminal design / production of the user. This module can be loaded directly into a variety of readers molds. Module uses voltage of 3.3V, through the SPI interface simple few lines can be directly connected to the user any CPU board communication module can guarantee stable and reliable work, reader distance.

Functions to interface RFID Reader:

1. RFID rfid(SS_PIN,RST_PIN) - To initialize the module and check the connection with arduino
2. RFID rfid.isCard() - To detect the presence of a rfid card trying to be scanned
3. RFID rfid.readCardSerial() - To detect the passing of data into the serial port

5.2 Software Section

5.2.1 MATLAB Data Acquisition and Analysis

The data obtained from the survey device is fed into an SD Card connected to the controller. The file is stored in a text format with the options selected from different users. The SD Card is inserted into the system and MathWorks MATLAB software is run. File reading commands are used to access the data in the file.

The number of questions and options are used to classify the data obtained. A counter is used to count the number of users opting for each option in a given question. Based on this a bar graph or pie chart is plotted to obtain graphical representation for analysis and interpretation. Further, based on the requirements and data, algorithms regarding data mining techniques are applied for efficient data analytics.

The flowchart in the figure 5.5 displays the complete information of data acquisition and the flow of data towards presentation and analysis. MATLAB code script does not perform any classification or prediction functions rather displays the detailed report of the data graphically.

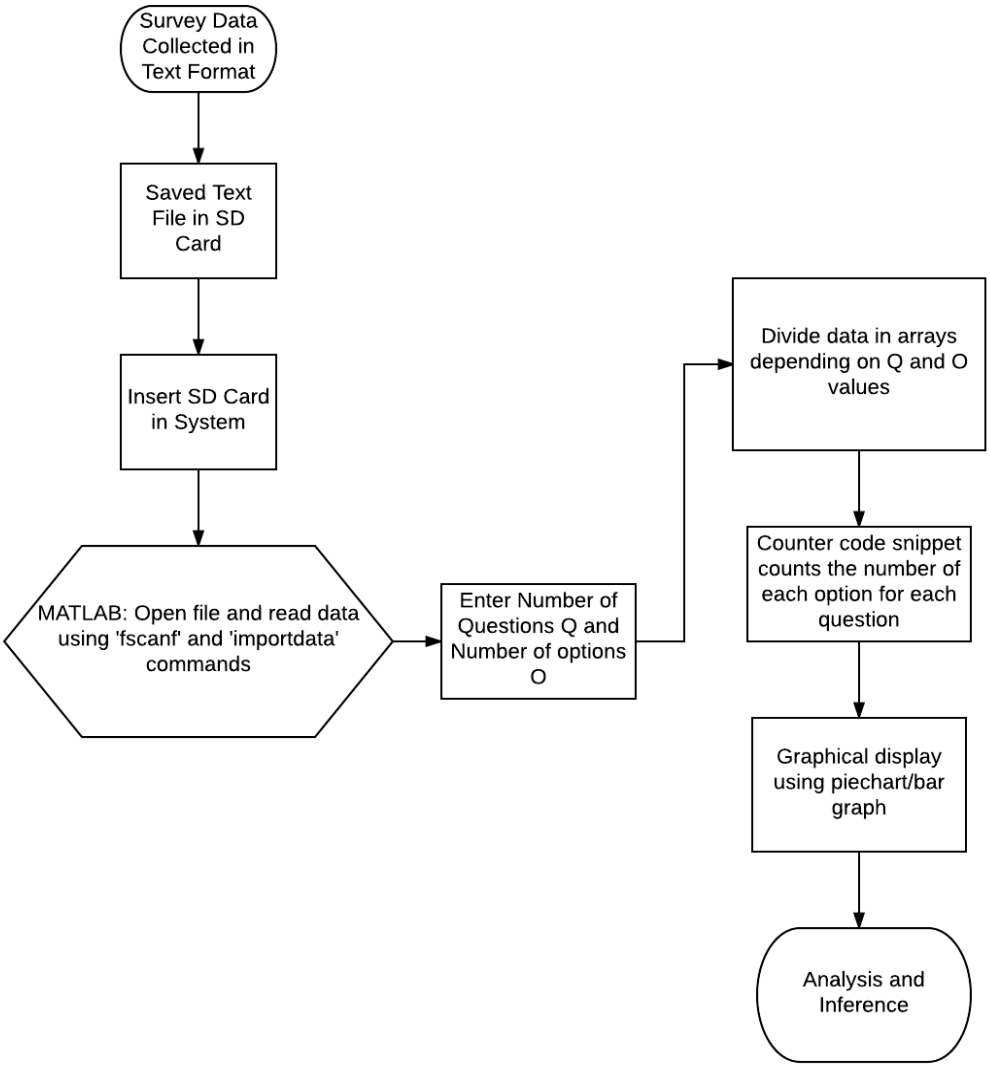

Figure 5.5: Flowchart representing the MATLAB interface for data interpretation

5.2.2 Scikit-learn Training and Prediction

Scikit-learn was used in python to implement the probabilistic model or the prediction of the output using the data generated by the survey system. The sklearn.datasets package embeds some small toy datasets as introduced in the Getting Started section. To evaluate the impact of the scale of the dataset (n_samples and n_features) while controlling the statistical properties of the data (typically the correlation and informativeness of the features), it is also possible to generate synthetic data.

Libraries like 'numpy' for standard loading of columnar data into numpy arrays and 'metrics' for developing the classification report were included. The python script created imports and reads the csv/txt file containing the data before further manipulation by sckit-learn.

Chapter 6

DATA ANALYTICS

6.1 Introduction to Data Analytics

Data analytics (DA) is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software. Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions and by scientists and researchers to verify or disprove scientific models, theories and hypotheses.

Data analytics can also be separated into quantitative data analysis and qualitative data analysis. The former involves analysis of numerical data with quantifiable variables that can be compared or measured statistically. The qualitative approach is more interpretive. At the application level, BI and reporting provides business executives and other corporate workers with actionable information about key performance indicators, business operations, customers and more.Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference consideration, complexity considerations, post-processing of discovered structures, visualization, and on-line updating.

More advanced types of data analytics include data mining, which involves sorting through large data sets to identify trends, patterns and relationships; predictive analytics, which seeks to predict customer behavior, equipment failures and other future events; and machine learning, an artificial intelligence technique that uses automated algorithms to churn through data sets more quickly than data scientists can do via conventional analytical modeling. Big data analytics applies data mining, predictive analytics and machine learning tools to sets of big data that often contain unstructured and semi-structured data. Text mining provides a means of analyzing documents, emails and other text-based content.

6.2 Machine Learning

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data.

The process of machine learning is similar to that of data mining. Both systems search through data to look for patterns. However, instead of extracting data for human comprehension, machine learning uses that data to detect patterns in data and adjust program actions accordingly. Machine learning algorithms are often categorized as being supervised or unsupervised. Supervised algorithms can apply what has been learned in the past to new data. Unsupervised algorithms can draw inferences from datasets.

Supervised Learning:

In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output. Supervised learning problems are categorized into ”regression” and ”classification” problems. In a regression problem, we are trying to predict results within a continuous output, meaning that we are trying to map input variables to some continuous function. In a classification problem, we are instead trying to predict results in a discrete output. In other words, we are trying to map input variables into discrete categories. Example:-

Given data about the size of houses on the real estate market, try to predict their price. Price as a function of size is a continuous output, so this is a regression problem.

Unsupervised Learning:

Unsupervised learning, on the other hand, allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables. We can derive this structure by clustering the data based on relationships among the variables in the data. With unsupervised learning there is no feedback based on the prediction results, i.e., there is no teacher to correct you. Example:-

Clustering: Take a collection of 1000 essays written on the US Economy, and find a way to automatically group these essays into a small number that are somehow similar or related by different variables, such as word frequency, sentence length, page count, and so on.

6.2.1 Logistic Regression

In statistics, logistic regression is a regression model where the dependent variable is categorical. This article covers the case of a binary dependent variablethat is, where it can take only two values, "0" and "1", which represent outcomes such as pass/fail, win/lose, alive/dead or healthy/sick. Cases where the dependent variable has more than two outcome categories may be analyzed in multinomial logistic regression, or, if the multiple categories are ordered, in ordinal logistic regression. In the terminology of economics, logistic regression is an example of a qualitative response/discrete choice model.

6.2.2 Mathematics involved in Logistic Regression

Logistic Regression is a classification type to address the binary classification problem in our project. The output 0 is taken as the 'positive class' and the output 1 is taken as the 'negative class'. The hypothesis should satisfy the criteria:
$0 \leq h_\theta(x) \leq 1$

The sigmoid function or logistic function is used which is represented by the following equations:

$h_\theta(x) = g(\theta^T x)$
$z = \theta^T x$
$g(z) = \frac{1}{1+e^{-z}}$
The function g(z), shown here, maps any real number to the (0, 1) interval, making it useful for transforming an arbitrary-valued function into a function better suited for classification.
For the parameter $_\theta$, the cost function is given below.

$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{Cost}(h_\theta(x^{(i)}), y^{(i)})$
$\text{Cost}(h_\theta(x), y) = -\log(h_\theta(x))$ if y=1
$\text{Cost}(h_\theta(x), y) = -\log(1h_\theta(x))$ if y=0
To optimize the training dataset and to minimize the error, the cost function is used in the gradient descent process.
Repeat{
$\theta_j := \theta_j - \alpha \frac{\delta}{\delta\theta_j} J(\theta)$
}

Chapter 7

RESULTS AND OBSERVATIONS

The micro-controller code generated was uploaded for different functional features like the LCD display, RFID security, button interface and the LED identifiers. Arduino Mega and Arduino Uno were uploaded with the code. The survey system hardware was now ready for conducting a survey. The survey is conducted after the admin initializes the survey using the RFID card. Each user presses the START button before submitting his choice. The data was then collected on an SD card interfaced to the Arduino Mega and a standalone system was used to perform MATLAB analytics and implementing machine learning algorithms using Scikit.

7.1 Generic Voting Results

The voting mode is where a single question is asked to the user to select a single option among many. After the survey was conducted and the SD card was analysed, the MATLAB script generated the following output for 8 options. The generic display of analytics is shown in the following figures. The bar graph shown in the figure is total count of each option. The pie chart depicts the result distribution of the option selections made along with the percent values.

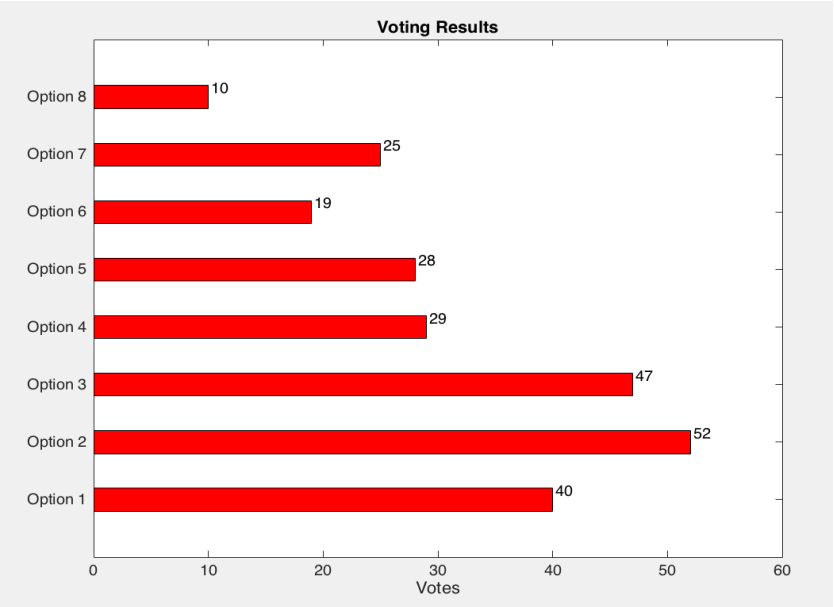

Figure 7.1: Generic voting data bar graph

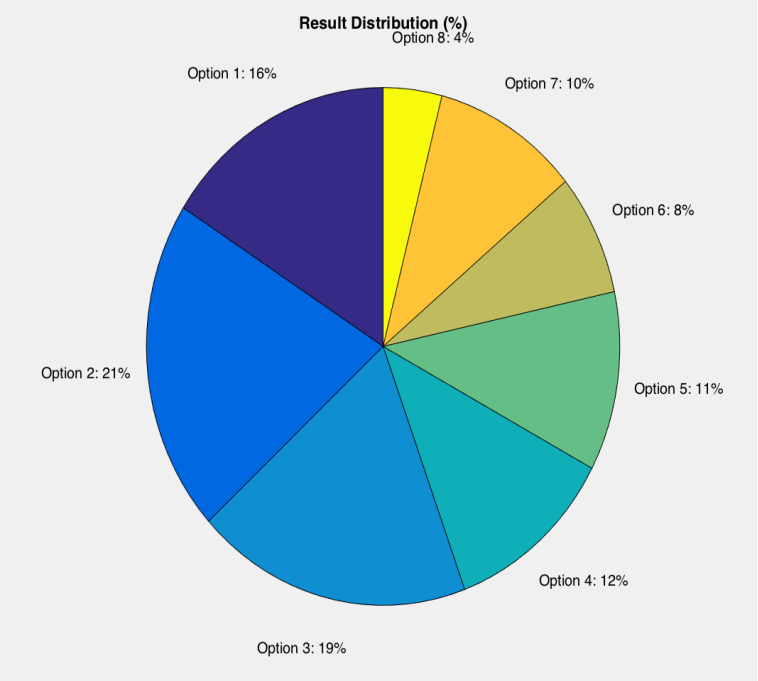

Figure 7.2: Generic voting data pie chart

7.2 Generic Polling Results

The polling mode is where multiple questions are asked to the user to decide the users opinion on a particular issue. Polling involves Yes/No type questions and the data is stored in the SD card in terms of bits 1/0 respectively. The survey is conducted with the same procedure and the results are analyzed using MATLAB.

The basic idea behind polling is to determine and analyze the parameters that influence the final decision of an issue. Four parameters are used in our model as prerequisite questions to the main question of the issue. The bar graph in the image show the summed up poll results parameter wise. The blue bar depicts the total end result for each parameter contributing to the affirmative end result, that is, bit 1. The yellow bar represents the total end result for each parameter contributing to the negative end result, that is, bit 0. The pie chart depicts the complete distribution in each row determines the distribution for each parameter. In case of parameter 1(P1), there are 55 percent users who are in affirmative to the main end question and in favour of the parameter 1 as well, while there are 25 percent users who answered yes for the main end question but not in favour of the parameter 1 question. Contrast to this there are 45 percent users who answered negative to the main end question while being in favour of the parameter 1 and 75 percent otherwise. This analysis is done similarly to all the parameters to get 8 pie charts in all.

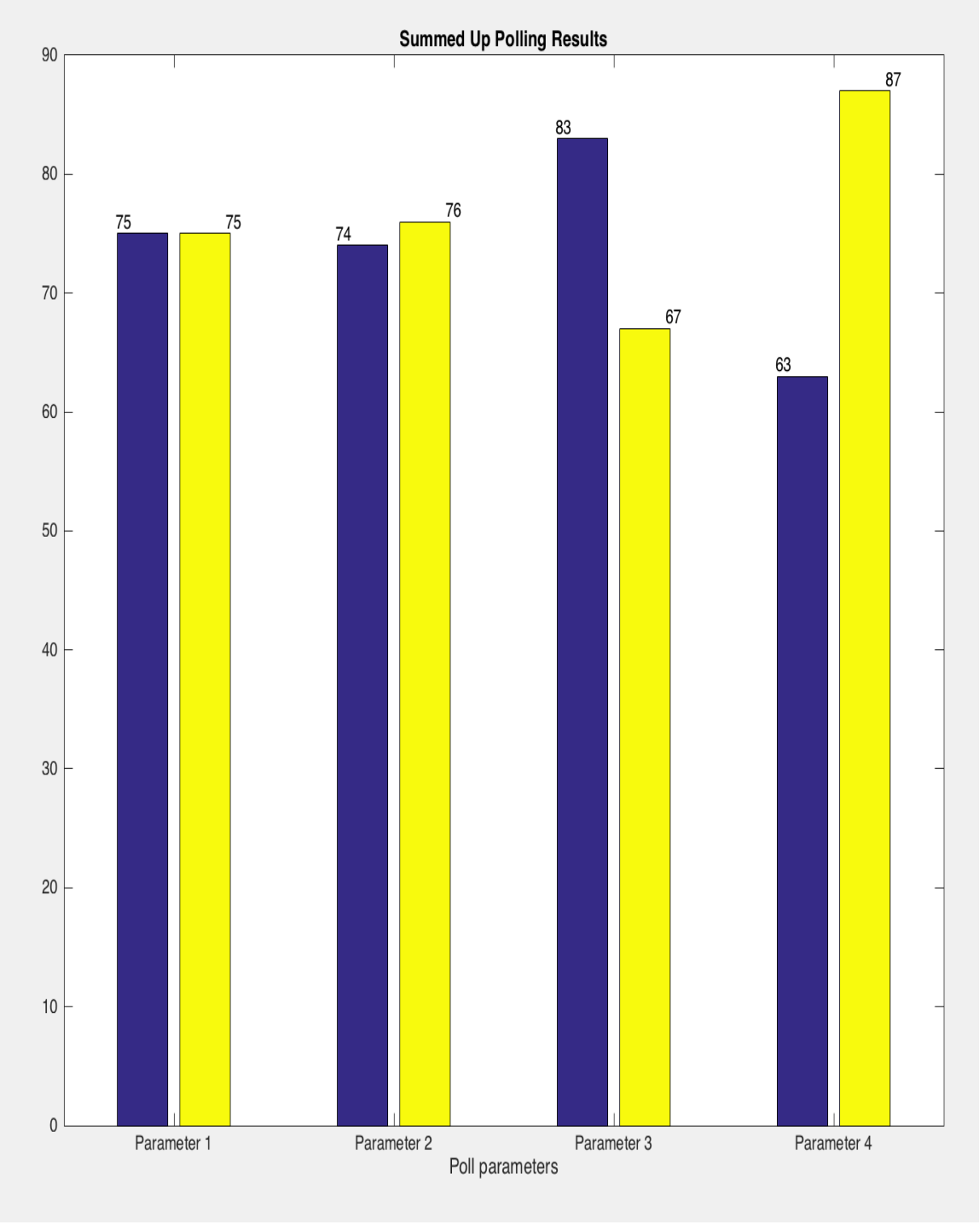

Figure 7.3: Generic polling data bar graph

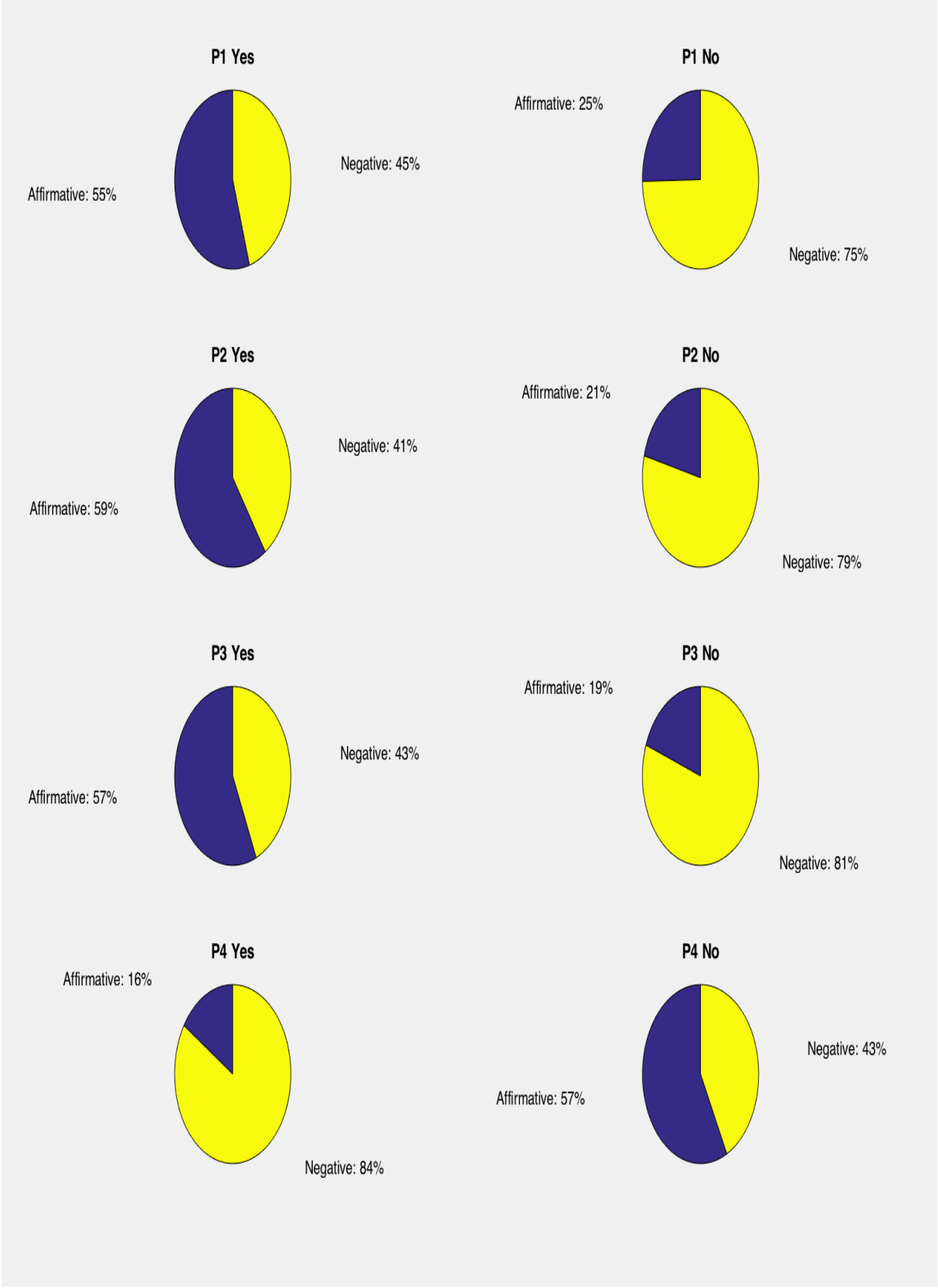

Figure 7.4: Generic polling data pie chart

7.3 Sample Surveys Conducted

The box prototype built in accordance to the design was used to conduct a few sample surveys. Since our project contains two different modes of operation- voting mode and polling mode, the sample surveys conducted includes both these modes.

7.3.1 Voting - Candidate Selection

The voting mode involves the answer of one question per user and is mainly based on selection of 1 option among many. In our prototype we have used 8 options. So the user selects 1 option among the 8 options given. The survey we conducted involved 8 people contesting similar to an election. The survey was done by 350 users. The bar graph displays the total counts for each person. From the output it is clear that Person 2 takes the lead with 63 votes and Person 3 running up with 54 votes. Pie chart depicts the result distribution with Person taking 18 percent share and Person 8 with 7 percent of the share. The figures below show the analysis output.

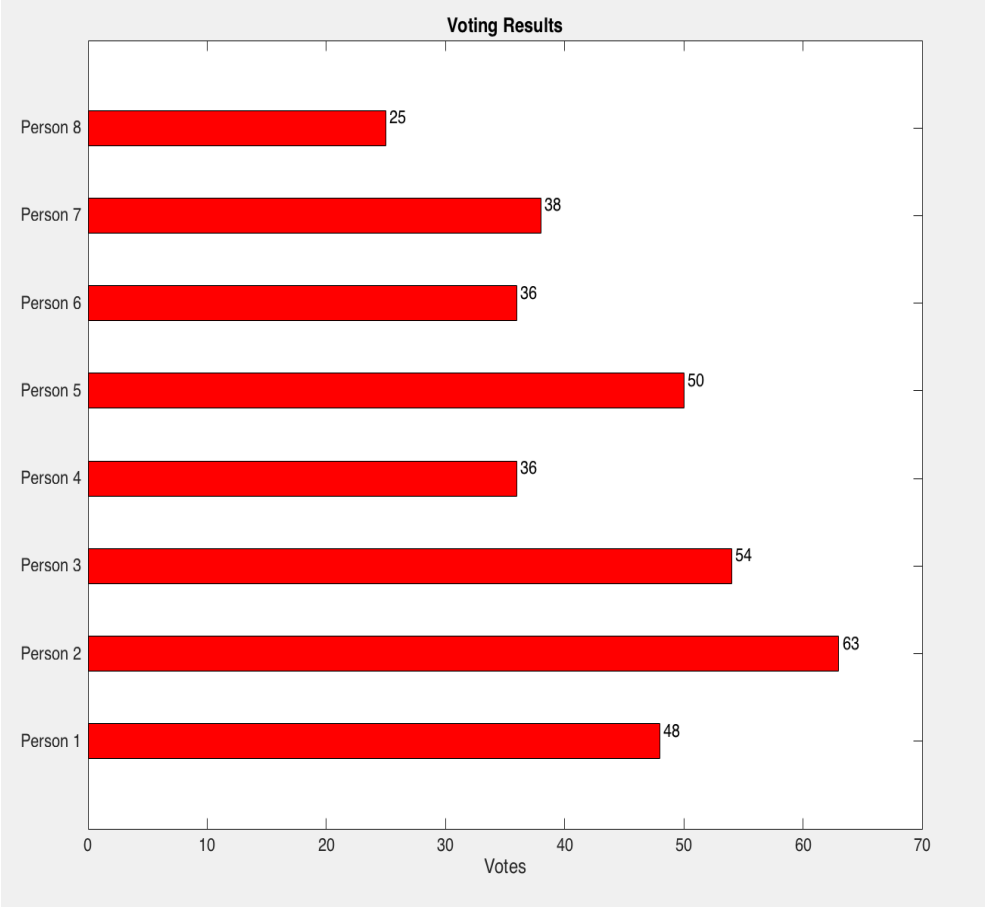

Figure 7.5: Candidate selection voting results bar graph

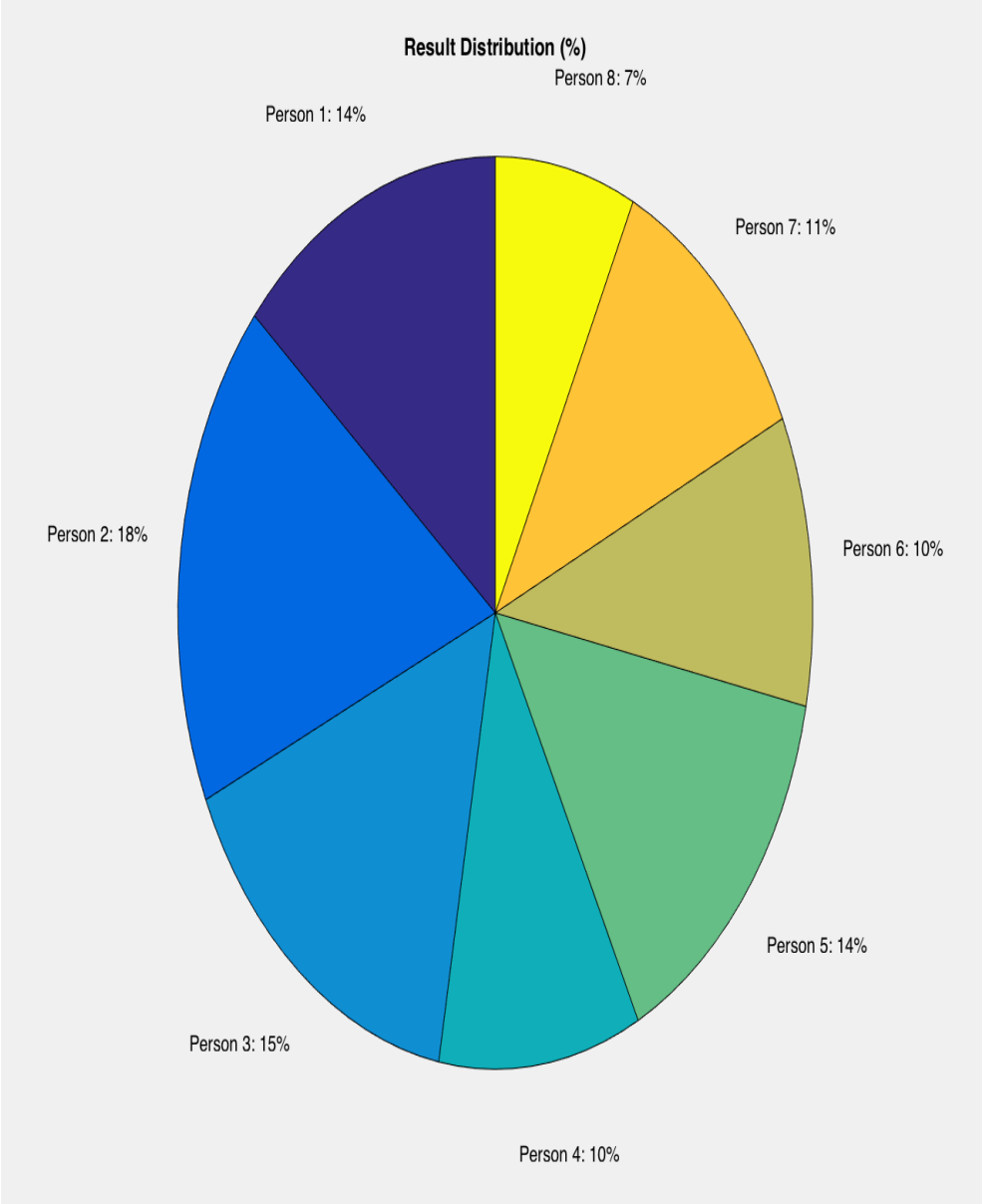

Figure 7.6: Candidate selection voting results pie chart

7.3.2 Polling - Student Smoking

The polling mode was used to address the issue of student smoking. The opinion of 150 users were taken. The data was then analyzed using a MATLAB script. The parameters used to generate the opinion of users include the gender, parents smoking, peer group smoking and the banning of smoking. These parameters at as the conditions to determine the final question whether the student smokes or not. The following figures were obtained as the output. The bar graph displays the total users answering in favour and against the parameter. The pie chart gives a detailed analysis output for the opinion. Each subplot corresponds to a parameter's distribution corresponding to the final result if the user smokes or not.

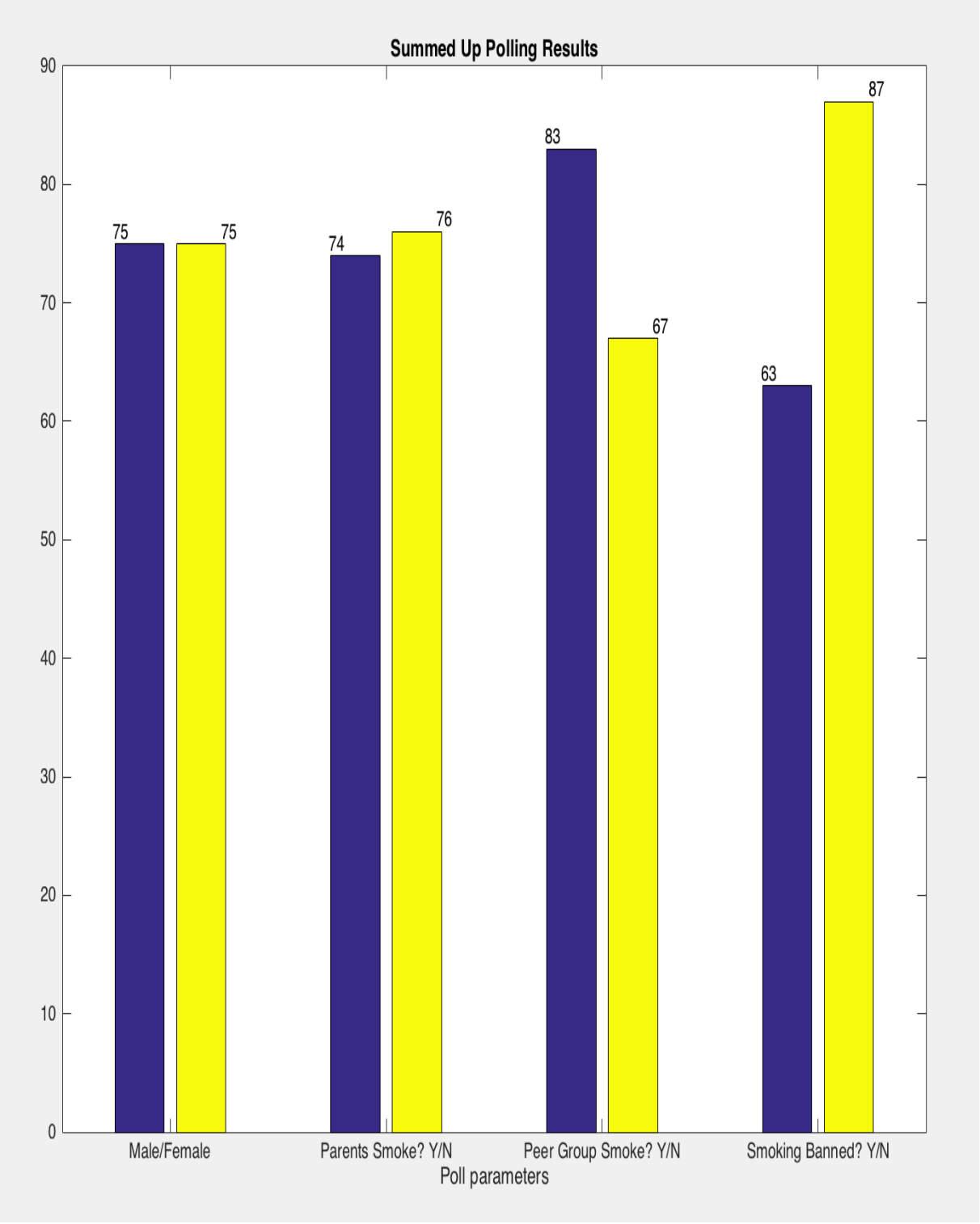

Figure 7.7: Student smoking polling results bar graph

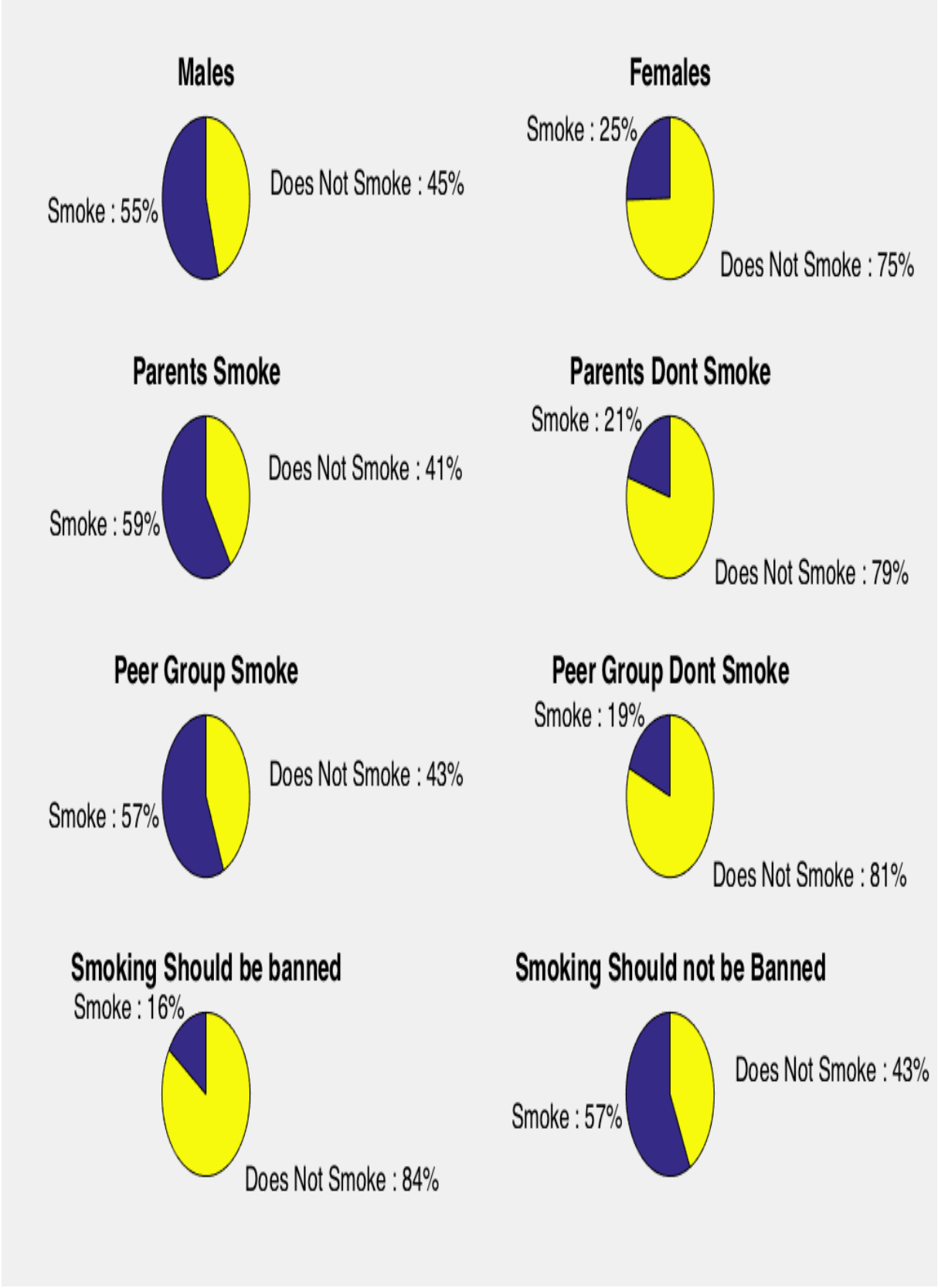

Figure 7.8: Student smoking polling results pie chart

7.3.3 Prediction

Prediction is an essential part in various fields of study which help preparing for future consequences. The data collected from the survey is subjected to machine learning algorithms to predict using a probabilistic approach as to what the output can be provided the same features.

Prediction can be useful in the following arenas:

1. Stock market share prices of companies in the near future
2. Opinion related prediction of a test user using polling mode
3. Politics and sports
4. Manufacturing industries

The implementation of machine learning algorithm multivariate linear regression to determine the probabilistic output is one of the major applications of the proposed system. The data obtained is in the .csv file form which is further imported by python script to perform the function.

The python script reads the file from the SD card reader and classifies the data as parameters for all of the columns except the last column of data which is the main end question containing the affirmative and negative class. This applies a multivariate regression classifier (logistic regression) on this data to analyze the probabilistic output for a test data. The input for the test data is taken when the python script is run and the user answers the same parameter questions displayed on the Terminal. The input is then used in the classifies as test data and the survey data taken as the training data.

The figure 7.9 displays the execution of the Scikit enabled python script for the above smoking polling example. The parameter questions are answered and the result displays whether the test user would end up smoking or not depending on the smoking data collected by the survey takers. The classification report is also displayed which shows the accuracy of the system at 91 percent based on f1-score. The confusion matrix displays the true positives, false negatives, false positives and true negatives. Precision and recall are calculated from this matrix. F1-score is the average of the precision and recall values.

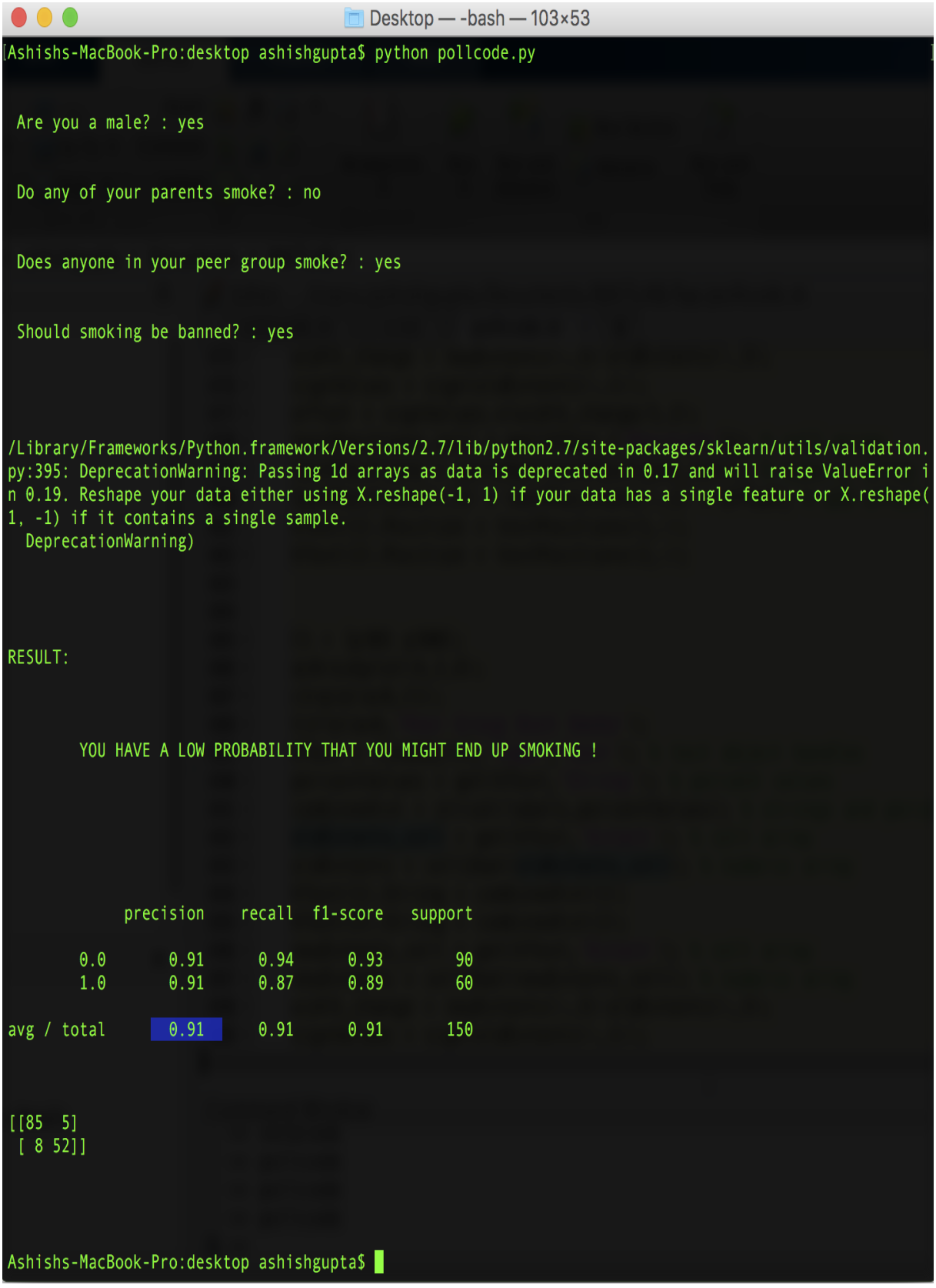

Figure 7.9: Prediction output and classification report

Chapter 8

CONCLUSION AND FUTURE SCOPE

8.1 Conclusion

The main perspective of this project was to develop a product similar to the EVMs currently in use, but with enhanced security, portability and most importantly, practicality. This was accomplished by making the product completely customizable, enabling the user to configure the device to meet any requirements. Data privacy and security was enhanced by many folds through data encryption and multi layer authentication. The polling scenario considered here returned high levels of accuracy and precision, which only increased with the size of the dataset processed.Device configured to register votes performed better, more efficiently than the existing purpose made devices. Analysis and representation of the data conveyed the progress and final outcome of the event in a simple and precise manner.

8.2 Future scope

Innovations are necessary to break the old practice to implement new technology with more advantages. Few more modules, newer and adaptable algorithms can be added to make the Modular survey system effective and with more security.

1. Data acquisition:

   The device can be easily configured to acquire repetitive data, by interfacing it with generic data gathering attachments such as a fingerprint sensor, camera module, and a keyboard. This same setup, for example, can be used to collect the identification data of an individual during Aadhaar enrolment or registration of new recruits in a company.

2. Security enhancement:

   Further improvement on the security end employs fingerprint authorization and encryption of the data stored locally. This two step authorization - RFiD and Biometric, ensures more security against data or device manipulation than any other commercial device.

3. Cloud integration:

   With the security offered by encryption, cloud integration opens up a new horizon of applications. Real time data transfer and analysis can now be done. Several modules can be deployed in an area linked through the cloud, with centralized controlling and analysis. Loss of data due to physical damage to the device or memory device corruption is also avoided through cloud backup. To further ease the process of data transfer and storing, data compression technique such as LZW is adaptable for our requirements.

4. Apriori Algorithm:

   Apriori is a classic algorithm in learning association rules in data mining. Apriori operates on databases having transactions (like details of a website frequentation, generic data from a poll, set of items bought by customers.With the use of a bottom up approach, frequent subsets are extended on item at a time (process known as candidate generation and these groups are tested against the data). Termination is when no successful extensions are found.

   Advantages of Apriori

   (a) Employs a large item set property

   (b) Easily parallelized

   (c) Easy to implement

   Disadvantages

   (a) Requires multiple database scans

   (b) Transaction database is assumed to be memory resident.

APPENDIX A

Arduino Mega-2560

Features:

- Operating Voltage:5V
- Input Voltage (recommended):7-12V
- Input Voltage (limit):6-20V
- Digital I/O Pins:54 (of which 15 provide PWM output)
- Analog Input Pins:16
- DC Current per I/O Pin:20 mA
- DC Current for 3.3V Pin:50 mA
- Flash Memory:256 KB of which 8 KB used by bootloader
- SRAM:8 KB
- EEPROM:4 KB
- Clock Speed 16 MHz
- Length: 101.52 mm
- Width: 53.3 mm
- Weight: 37 g

Description:

The Arduino Mega 2560 is a microcontroller board based on the ATmega2560. It has 54 digital input/output pins (of which 14 can be used as PWM outputs), 16 analog inputs, 4 UARTs(hardware serial ports), a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header,and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started.

APPENDIX B

Arduino Uno

Features:

-Microcontroller: ATmega328P

-Operating Voltage:5V

-Input Voltage (recommended): 7-12V

-Input Voltage (limit): 6-20V

-Digital I/O Pins: 14 (of which 6 provide PWM output)

-PWM Digital I/O Pins: 6

-Analog Input Pins: 6

-DC Current per I/O Pin: 20 mA

-DC Current for 3.3V Pin: 50 mA

-Flash Memory: 32 KB (ATmega328) which 0.5 KB used by bootloader

-SRAM: 2 KB (ATmega328P)

-EEPROM: 1 KB (ATmega328P)

-Clock Speed: 16 MHz

-Length: 68.6 mm

-Width: 53.4 mm

-Weight: 25 g

Description:

The Arduino Uno is a microcontroller board based on the ATmega328 (datasheet). It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header, and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started.

APPENDIX C

LCD 16x2

Features:

- 5 x 8 dots with cursor
- Built-in controller (KS 0066 or Equivalent)
- + 5V power supply (Also available for + 3V)
- 1/16 duty cycle
- B/L to be driven by pin 1, pin 2 or pin 15, pin 16 or A.K (LED)
- N.V. optional for + 3V power supply

Description:

LCD (Liquid Crystal Display) screen is an electronic display module and find a wide range of applications. A 16x2 LCD display is very basic module and is very commonly used in various devices and circuits. These modules are preferred over seven segments and other multi segment LEDs. The reasons being: LCDs are economical; easily programmable; have no limitation of displaying special and even custom characters (unlike in seven segments), animations and so on.
A 16x2 LCD means it can display 16 characters per line and there are 2 such lines. In this LCD each character is displayed in 5x7 pixel matrix. This LCD has two registers, namely, Command and Data.The command register stores the command instructions given to the LCD. A command is an instruction given to LCD to do a predefined task like initializing it, clearing its screen, setting the cursor position, controlling display etc. The data register stores the data to be displayed on the LCD. The data is the ASCII value of the character to be displayed on the LCD.

APPENDIX D

RFID Module MFRC522

Features:

-Current :13-26mA / DC 3.3V

-Idle Current :10-13mA / DC 3.3V

-Sleep Current¡80uA

-Peak Current¡30mA

-Operating Frequency: 13.56MHz¡30mA

-1 Piece new MFRC-522-RC522-RFID-RF-IC-card-reader-sensor-module

-Module can be used with Arduino, Mbed, ARM or any other MCU

Description:

The MFRC522 is a highly integrated reader/writer IC for contactless communication at 13.56 MHz. The MFRC522 reader supports ISO/IEC 14443 A/MIFARE and NTAG. The MFRC522s internal transmitter is able to drive a reader/writer antenna designed to communicate with ISO/IEC 14443 A/MIFARE cards and transponders without additional active circuitry. The receiver module provides a robust and efficient implementation for demodulating and decoding signals from ISO/IEC 14443 A/MIFARE compatible cards and transponders. The digital module manages the complete ISO/IEC 14443 A framing and error detection (parity and CRC) functionality.

The MFRC522 supports MF1xxS20, MF1xxS70 and MF1xxS50 products. The MFRC522 supports contactless communication and uses MIFARE higher transfer speeds up to 848 kBd in both directions.

APPENDIX E

Micro SD card adapter

Features:

-The level conversion circuit board that can interface level is 5V or 3.3VSupport Micro SD Card, Micro SDHC card (high-speed card).

-Power supply is 4.5V 5.5V, 3.3V voltage regulator circuit board; Communication interface is a standard SPI interface.

-Level conversion circuit: Micro SD card into the direction of signals into 3.3V, MicroSD card toward the direction of the control interface MISO signal is also converted to 3.3V, general AVR microcontroller system can read the signal.

-Power supply: 4.5V - 5.5V

-Control Interface: GND, VCC, MISO, MOSI, SCK, CS

-Size: 45 28 mm

-Net weight: 6g

Description:

Support Micro SD Card, Micro SDHC card (high-speed card) The level conversioncircuit board that can interface level is 5V or 3.3V. Communication interface is a standard SPI interface Control Interface: A total of six pins (GND, VCC, MISO, MOSI, SCK, CS), GND to ground, VCC is the power supply, MISO, MOSI, SCK is the SPI bus, CS is the chip select signal pin; 3.3V regulator circuit: LDO regulator output 3.3V as level converter chip, Micro SD card supply; Level conversion circuit: Micro SD card into the direction of signals into 3.3V, MicroSD card toward the direction of the control interface MISO signal is also converted to 3.3V, general AVR microcontroller system can read the signal.

REFERENCES

[1] Joao Tiago and Medeiros Paulo (2009) - "Efficient Storage of Data in Cloud Computing"

[2] Vaidy Krishnan (2013) - "Solving the Internet of Things Last-Mile Problem"

[3] Hyun-Suk Yu, Yvette E. Gelogo, Kyung Jung Kim - "Securing Data Storage in Cloud Computing"

[4] Ben James - "5 Steps to Self-Service Analytics that Scales"

[5] Jeff Mills - "How to Build a Culture of Self-Service Analytics"