Training Report

Bachelor of Technology
(Information Technology)

Submitted By :
Gurnoor Singh
(1411254)

under the guidance of
Er. Sachin Bagga
(Assistant Professor, IT Department)

Department of Information Technology
Guru Nanak Dev Engineering College
Ludhiana 141006

CONTENTS

INTRODUCTION TO BIG DATA

Big data is a term that describes the large volume of data both structured and unstructured that inundates a business on a day-to-day basis. But its not the amount of data thats important. Its what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

1.1 Big Data Characteristics

Big data can be described by the following characteristics:

1. Volume The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can actually be considered big data or not.

2. Variety The type and nature of the data. This helps people who analyze it to effectively use the resulting insight.

3. Velocity In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.

4. Variability Inconsistency of the data set can hamper processes to handle and manage it.

5. Complexity Today's data comes from multiple sources, which makes it difficult to link, match, cleanse and transform data across systems. However, its necessary to connect and correlate relationships, hierarchies and multiple data linkages or your data can quickly spiral out of control.

1.2 Importance of Big Data

The importance of big data doesnt revolve around how much data you have, but what you do with it. You can take data from any source and analyze it to find answers that enable

- Cost reductions
- Time reductions

- New product development and optimized offerings

- Smart decision making.

When you combine big data with high-powered analytics, you can accomplish business-related tasks such as:

- Determining root causes of failures, issues and defects in near-real time.

- Generating coupons at the point of sale based on the customers buying habits.

- Recalculating entire risk portfolios in minutes.

- Detecting fraudulent behaviour before it affects your organization.

1.3   Uses of Big Data

Big data affects organizations across practically every industry. See how each industry can benefit from this onslaught of information.

- Banking

  With large amounts of information streaming in from countless sources, banks are faced with finding new and innovative ways to manage big data. While its important to understand customers and boost their satisfaction, its equally important to minimize risk and fraud while maintaining regulatory compliance. Big data brings big insights, but it also requires financial institutions to stay one step ahead of the game with advanced analytics.

- Education

  Educators armed with data-driven insight can make a significant impact on school systems, students and curriculums. By analyzing big data, they can identify at-risk students, make sure students are making adequate progress, and can implement a better system for evaluation and support of teachers and principals.

- Government

  When government agencies are able to harness and apply analytics to their big data, they gain significant ground when it comes to managing utilities, running agencies, dealing with traffic congestion or preventing crime. But while there are many advantages to big data, governments must also address issues of transparency and privacy.

CHAPTER 2

INTRODUCTION TO DATA MINING

Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of 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 considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

2.1 Data Mining Parameters

In data mining, association rules are created by analyzing data for frequent if/then patterns, then using the support and confidence criteria to locate the most important relationships within the data. Support is how frequently the items appear in the database, while confidence is the number of times if/then statements are accurate. Other data mining parameters include Sequence or Path Analysis, Classification, Clustering and Forecasting. Sequence or Path Analysis parameters look for patterns where one event leads to another later event. A Sequence is an ordered list of sets of items, and it is a common type of data structure found in many databases. A Classification parameter looks for new patterns, and might result in a change in the way the data is organized. Classification algorithms predict variables based on other factors within the database.

2.2 Benefits Of Data Mining

In general, the benefits of data mining are -

- The ability to uncover hidden patterns

- Uncover relationships in data

- Can be used to make predictions that impact businesses.

- Sales and marketing departments can mine customer data.

3

CHAPTER 3

PHP

PHP is a server-side scripting language designed primarily for web development but also used as a general-purpose programming language. Originally created by Rasmus Lerdorf in 1994, the PHP reference implementation is now produced by The PHP Development Team. PHP originally stood for Personal Home Page, but it now stands for the recursive acronym PHP: Hypertext Pre-processor.

Figure 3.1: PHP logo

PHP code may be embedded into HTML or HTML5 mark-up, or it can be used in combination with various web template systems, web content management systems and web frameworks. PHP code is usually processed by a PHP interpreter implemented as a module in the web server or as a Common Gateway Interface (CGI) executable. The web server software combines the results of the interpreted and executed PHP code, which may be any type of data, including images, with the generated web page. PHP code may also be executed with a command-line interface (CLI) and can be used to implement standalone graphical applications. The standard PHP interpreter, powered by the Zend Engine, is free software released under the PHP License. PHP has been widely ported and can be deployed on most web servers on almost every operating system and platform, free of charge. The PHP language evolved without a written formal specification or standard until 2014, leaving the canonical PHP interpreter as a de facto standard. Since 2014 work has gone on to create a formal PHP specification.

CHAPTER 4

HUMAN PSYCHOLOGY

Psychology is the science of behaviour and mind, embracing all aspects of conscious and unconscious experience as well as thought. It is an academic discipline and a social science which seeks to understand individuals and groups by establishing general principles and researching specific cases. In this field, a professional practitioner or researcher is called a psychologist and can be classified as a social, behavioural, or cognitive scientist. Psychologists attempt to understand the role of mental functions in individual and social behaviour, while also exploring the physiological and biological processes that underlie cognitive functions and behaviours. Psychologists explore behaviour and mental processes, including perception, cognition, attention, emotion (affect), intelligence, phenomenology, motivation (conation), brain functioning, and personality. This extends to interaction between people, such as interpersonal relationships, including psychological resilience, family resilience, and other areas. Psychologists of diverse orientations also consider the unconscious mind. Humanistic psychology is a psychological perspective that emphasizes the study of the whole person. Humanistic psychologists look at human behaviour not only through the eyes of the observer, but through the eyes of the person doing the behaving. Humanistic psychologists believe that an individual's behaviour is connected to his inner feelings and self-image. Unlike the behaviourists, humanistic psychologists believe that humans are not solely the product of their environment. Rather humanistic psychologists study human meanings, understandings, and experiences involved in growing, teaching, and learning. They emphasize characteristics that are shared by all human beings such as love, grief, caring, and self-worth. Humanistic psychologists study how people are influenced by their self-perceptions and the personal meanings attached to their experiences. Humanistic psychologists are not primarily concerned with instinctual drives, responses to external stimuli, or past experiences. Rather, they consider conscious choices, responses to internal needs, and current circumstances to be important in shaping human behaviour.

CHAPTER 5

COMPUTATIONAL BIOLOGY

Computational biology involves the development and application of data-analytical and theoretical methods, mathematical modelling and computational simulation techniques to the study of biological, behavioural, and social systems. The field is broadly defined and includes foundations in computer science, applied mathematics, animation, statistics, biochemistry, chemistry, biophysics, molecular biology, genetics, genomics, ecology, evolution, anatomy, neuroscience, and visualization. Computational biology is different from biological computation, which is a subfield of computer science and computer engineering using bioengineering and biology to build computers, but is similar to bioinformatics, which is an interdisciplinary science using computers to store and process biological data. Computational Biology, which includes many aspects of bioinformatics, is the science of using biological data to develop algorithms or models to understand among various biological systems and relationships. Until recently, biologists did not have access to very large amounts of data which have become commonplace, particularly in molecular biology and genomics. Researchers were able to develop analytical methods for interpreting biological information, but were unable to share them quickly among colleagues.

5.1   Bioinformatics

Bioinformatics is the use of computers for the acquisition, management, and analysis of biological information. It incorporates elements of molecular biology, computational biology, database computing, and the Internet. Bioinformatics is clearly a multi-disciplinary field including: computer systems management networking, database design, computer programming, and molecular biology. Bioinformatics is both an umbrella term for the body of biological studies that use computer programming as part of their methodology, as well as a reference to specific analysis "pipelines" that are repeatedly used, particularly in the field of genomics. Common uses of bioinformatics include the identification of candidate genes and single nucleotide polymorphisms (SNPs). Often, such identification is made with the aim of better understanding the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), or differences between populations. In a less formal way, bioinformatics also tries to understand the organisational principles within nucleic acid and protein sequences, called proteomics.
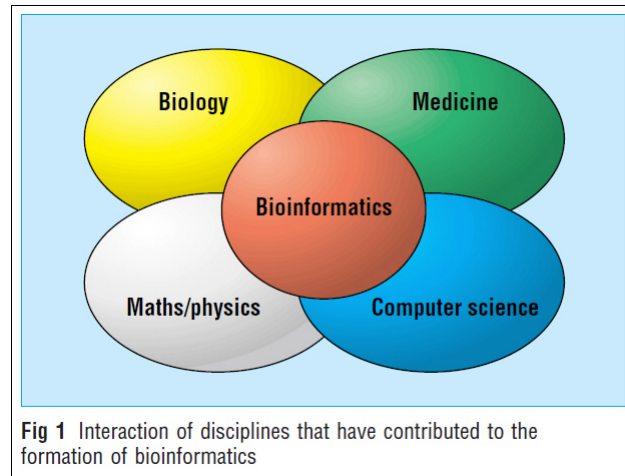
Fig 1 Interaction of disciplines that have contributed to the formation of bioinformatics

Figure 5.1: Bioinformatics

5.2 Sequence Analysis

1. DNA SEQUENCING

   Before sequences can be analyzed they have to be obtained. DNA sequencing is still a nontrivial problem as the raw data may be noisy or afflicted by weak signals. Algorithms have been developed for base calling for the various experimental approaches to DNA sequencing.

2. SEQUENCE ASSEMBLY

   Most DNA sequencing techniques produce short fragments of sequence that need to be assembled to obtain complete gene or genome sequences. The so-called shotgun sequencing technique generates the sequences of many thousands of small DNA fragments

3. GENOME ANNOTATION

   In the context of genomics, annotation is the process of marking the genes and other biological features in a DNA sequence. This process needs to be automated because most genomes are too large to annotate by hand, not to mention the desire to annotate as many genomes as possible, as the rate of sequencing has ceased to pose a bottleneck.

JAVA

Java is a general-purpose computer programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA),[15] meaning that compiled Java code can run on all platforms that support Java without the need for recompilation. Java applications are typically compiled to byte code that can run on any Java virtual machine (JVM) regardless of computer architecture. As of 2016, Java is one of the most popular programming languages in use, particularly for client-server web applications, with a reported 9 million developers. Java was originally developed by James Gosling at Sun Microsystems (which has since been acquired by Oracle Corporation) and released in 1995 as a core component of Sun Microsystems' Java platform. The language derives much of its syntax from C and C++, but it has fewer low-level facilities than either of them.

Figure 6.1: Java Logo

The original and reference implementation Java compilers, virtual machines, and class libraries were originally released by Sun under proprietary licenses. As of May 2007, in compliance with the specifications of the Java Community Process, Sun relicensed most of its Java technologies under the GNU General Public License. Others have also developed alternative implementations of these Sun technologies, such as the GNU Compiler for Java (byte code compiler), GNU Class path (standard libraries), and IcedTea-Web (browser plug-in for applets). The latest version is Java 8 which is the only version currently supported for free by Oracle, although earlier versions

are supported both by Oracle and other companies on a commercial basis.

6.1 Java Platform

One design goal of Java is portability, which means that programs written for the Java platform must run similarly on any combination of hardware and operating system with adequate runtime support. This is achieved by compiling the Java language code to an intermediate representation called Java byte code, instead of directly to architecture-specific machine code. Java byte code instructions are analogous to machine code, but they are intended to be executed by a virtual machine (VM) written specifically for the host hardware. End users commonly use a Java Runtime Environment (JRE) installed on their own machine for standalone Java applications, or in a web browser for Java applets. Standard libraries provide a generic way to access host-specific features such as graphics, threading, and networking. The use of universal bytecode makes porting simple. However, the overhead of interpreting bytecode into machine instructions made interpreted programs almost always run more slowly than native executables. Just-in-time (JIT) compilers that compile byte codes to machine code during runtime were introduced from an early stage. Java itself is platform-independent and is adapted to the particular platform it is to run on by a Java virtual machine for it, which translates the Java byte code into the platform's machine language.

6.2 Automatic Memory Management

Java uses an automatic garbage collector to manage memory in the object lifecycle. The programmer determines when objects are created, and the Java runtime is responsible for recovering the memory once objects are no longer in use. Once no references to an object remain, the unreachable memory becomes eligible to be freed automatically by the garbage collector. Something similar to a memory leak may still occur if a programmer's code holds a reference to an object that is no longer needed, typically when objects that are no longer needed are stored in containers that are still in use. If methods for a nonexistent object are called, a "null pointer exception" is thrown. One of the ideas behind Java's automatic memory management model is that programmers can be spared the burden of having to perform manual memory management. In some languages, memory for the creation of objects is implicitly allocated on the stack or explicitly allocated and deallocated from the heap. In the latter case, the responsibility of managing memory resides with the programmer.

CHAPTER 7

SERVER ADMINISTRATION

Server Administration is an advanced computer networking topic that includes server installation and configuration, server roles, storage, Active Directory and Group Policy, file, print, and web services, remote access, virtualization, application servers, troubleshooting, performance, and reliability. This course is comprised of 15 lessons that use Windows Server to study and experiment with server administration. Each lesson includes a combination of Wikipedia and Microsoft readings, YouTube videos, and hands-on learning activities. The course also assists learners in preparing for the Microsoft MTA Exam 98-365: Windows Server Administration Fundamentals. This entire Wikiversity course can be downloaded in book form by selecting Download Learning Guide in the sidebar. The corresponding Wikipedia reading collection can be downloaded in book form by selecting Download Reading Guide.

7.1   Server Administrator

A server administrator or admin has the overall control of a server. This is usually in the context of a business organization, where a server administrator oversees the performance and condition of multiple servers in the business organization, or it can be in the context of a single person running a game server. The Server Administrator's role is to design, install, administer, and optimize company servers and related components to achieve high performance of the various business functions supported by the servers as necessary. This includes ensuring the availability of client/server applications, configuring all new implementations, and developing processes and procedures for ongoing management of the server environment. Where applicable, the Server Administrator will assist in overseeing the physical security, integrity, and safety of the data centre/server farm.

7.2   System Administrator Duties And Skills

Due to the wide range of job responsibilities for system administrators in various organizations, system administrators' job skill requirements are often broad, as are salary ranges. In general, sysadmins must be comfortable working with application and file servers, desktops, networks, databases, information security systems and storage. Familiarity with multiple operating systems, as well as scripting and programming, is often required. Increasingly, virtualization and cloud

computing skills have also become essential to the job. Because tasks generally include provisioning, configuring and managing physical and virtual servers, as well as the software that runs on the servers and the hardware that supports them, a system administrator should feel comfortable installing and troubleshooting IT resources, establishing and managing user accounts, upgrading and patching software, and performing backup and recovery tasks. Nontechnical skills are equally important for sysadmins. Because the system administrator interacts with people in so many areas of IT and business, soft skills (people skills) are just as necessary as hard skills. When IT services are slow or down entirely, a system administrator must be able to work under pressure, read a situation as it unfolds and quickly decide upon a response that yields the best result for all involved.

7.3 IT System Administrator Certifications

System administrators are expected to have at least one, but preferably multiple, certifications for the job. Depending on the technologies used within an enterprise, common certifications in demand include Microsoft Certified Solutions Associate (MCSA), CompTIA Server+, Cisco Certified Network Associate (CCNA) and Red Hat Certified System Administrator (RHCSA).

CHAPTER 8

HUMAN VALUES

Human values have been employed in so distinctively different ways in human discourse. It is often said that a person has a value or an object has a value. These two usages have been explicitly recognised by writers in various disciplines such as Charles Mortris in Philosophy, Brevster Smith in Psychology and Roibin Williams in Sociology. If one wants to know the origin of the term VALUE, it may be stated very firmly that the term VALUE comes from the Latin word VALERE which means to be of worth. Whereas, the concise Oxford Dictionary defines the term VALUE as the worth, desirability or utility of a thing.

In fact, it is difficult to define values, for they are as comprehensive in a nature as our human life. Somewhere, some other dictionary states that Value is that which renders anything useful, worthy or estimable. It is price, worth or importance of a thing.A Value is a concept explicit of implicit, distinctive of an individual or characteristics of a group of those desirable traits which influence the selection from available modes and ends of action.

In fact, value is an abstract term which is commonly regarded as an economic conception. In the words of John Dewey, Value means primarily, to price, to esteem, to appraise, to estimate. It means the act of cherishing something holding it clear and also, the act of passing judgement upon the nature and amount of its value as compared with something else, According to Rokeach, Value is an enduring belief, a specific mode of conduct or an end state of existence, along a continuum of relative importance. Values are part and parcel of philosophy. Hence, aims of education are naturally concerned with values. Ail education is, in fact, very naturally value-oriented. Each educational goal, whether originating in a person, a family, a community, a school or an educational system, is believed to be good. Good is intended to mean here avoidance of bad. If possible objective, is not good, then there is no reason for pursuing. But again, the same question spurts out, and when the question what is a value? spurts out, we know something of its religion, philosophy and ideology. The guiding social aims and beliefs which are regarded as the important aspects of a culture, then, the different aspects of culture are also valued by the people; and the ideas lying behind which they think worthwhile, are called as VALUES!

Values are defined as something which are desirable and worthy of esteem for their own sake. Human values are defined as those values which help man to live in harmony with the world. Values that may be included in the general definition of human values are love, brotherhood, respect for others including plants and animals honesty, sincerity, truthfulness, nonviolence, gratitude, tolerance, a sense of responsibility, cooperation, self-reliance, secularism and internationalism.

CHAPTER 9

PERFORMANCE ENHANCEMENT

Computer Performance is the amount of work accomplished by a computer system. Computer software performance, particularly software application response time, is an aspect of software quality that is important in humancomputer interactions. Depending on the context, high computer performance may involve one or more of the following:

- Short response time for a given piece of work

- High throughput (rate of processing work)

- Low utilization of computing resource(s)

- High availability of the computing system or application

- Fast (or highly compact) data compression and decompression

- High bandwidth

- Short data transmission time

Memory Hierarchy is important for Performance Enhancement as it ensures that maximum elements can be processed with minimum penalty of computer cycles. In computer science, locality
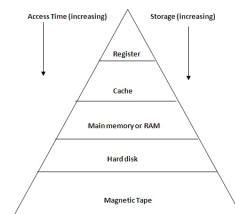
Figure 9.1: Performance Enhancement

of reference, also known as the principle of locality, is a term for the phenomenon in which the same values, or related storage locations, are frequently accessed, depending on the memory access pattern. There are two basic types of reference locality temporal and spatial locality. Temporal

locality refers to the reuse of specific data, and/or resources, within a relatively small time duration. Spatial locality refers to the use of data elements within relatively close storage locations. Sequential locality, a special case of spatial locality, occurs when data elements are arranged and accessed linearly, such as, traversing the elements in a one-dimensional array.

INTRODUCTION TO RASPBERRY PI

Raspberry pi is as such most inspiring computer available today because most devices that we use today like phones , tablets are designed in such a way that that we cannot manipulate them , From manipulation i mean we cannot design it in such a way to get a specific desired output , But raspberry is exactly opposite. It was invented in the UK as a device for promoting the teaching of computer science and coding, and its development is overseen by The Raspberry Pi Foundation charity. The first-generation Raspberry Pi appeared in 2012 and the latest Raspberry Pi 3.0 arrived in February this year. Eight million devices have been sold worldwide, the Pi Foundation says. (Source Google). It is credit-card sized computer which can be plugged into your TV and a
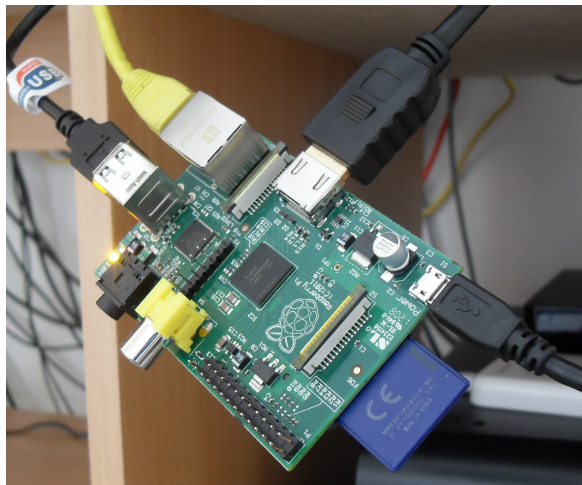

Figure 10.1: Raspberry pi logo

keyboard , which can be used for many of the things that an average computer does like watching movies , word processing, playing light games etc.

10.1   Comparing Raspberry pi with PC

So if we compare both , In Raspberry pi we can only download open-source operating system and the Ram , Graphics and processor of Raspberry pi is comparatively low as compare to that of

Compare Raspberry Pi and a PC

Components	Laptop or PC	Raspberry Pi Ver B
Processor	Intel 2.2 Ghz,Dual Core	700 Mhz,Single Core Arm 11
RAM	6GB	512 MB
Graphics	Intel HD 3000	Dual core video core IV
Ethernet	Yes	Yes
USB 2.0	Yes	Yes
Video O/P	VGA ,HDMI	Composite RCA HDMI
Audio O/P	Yes	Yes
Storage	500 GB Harddisk	32 GB SD Card
Operating System	Linux/Windows	Only Linux
Dimensions	14 inch laptop	8.6x5.4x1.7 cm

Figure 10.2: Raspberry pi logo

personal computer which works completely fine for doing basic operations other than playing high end games like GTA V.
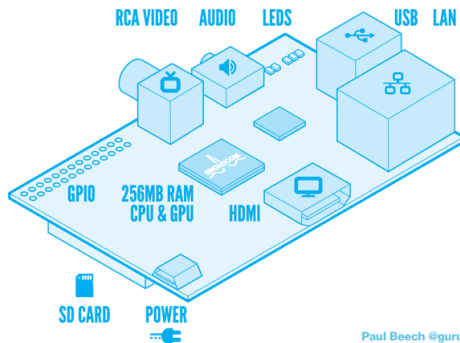
10.2   Basics of Raspberry pi

Figure 10.3: Raspberry pi

We have HDMI port to connect raspberry pi it to monitor . We have a dedicated SD card slot for installing open-source operating system. Here we have blue colored SD card in the picture, (in the latest version) integrated wifi and Bluetooth. We also have a camera for recording and a headphone jack using which we can listen to songs and it can also be useful for any automation project. We also have a 26 pin GPIO (General Purpose I/O) which is used for carrying out specific job from Raspberry Pie.

10.3   Comparing Raspberry pie Models

There are basically two models

- Model A

- Model B

	Model A	Model B
RRP	$25	$35
System on a Chip	Broadcom BCM2835	
CPU	700 MHz ARM1176JZF-S core	
GPU	Broadcom VideoCore IV, OpenGL ES 2.0. Device capable of MPEG-2 and VC-1, 1080p30 h.264/MPEG-4 AVC decoding and encoding.	
Memory (SDRAM)	256 MB, shared with GPU	512 MB (models build since October 15th 2012), shared with GPU
USB 2.0	1	2 (integrated USB hub)
Video Out	Composite RCA (PAL and NTSC), HDMI (also Display Serial Interface for LCD panels)	
Audio Out	3.5 mm jack, HDMI	
Storage	SD/MMC/SDIO card slot	
Network	No connector	RJ45 Ethernet through integrated USB hub
Peripheral connectors	8 x GPIO, UART, I²C bus, SPI bus	
Power rating	300 mA (1.5 W)	700 mA (3.5 W)
Power source	5 volt via MicroUSB or GPIO header	

Figure 10.4: Comparing Raspberry pie Models

It cost around $25(1750) for Model A and $35 (2450) for Model B . In Model A we dont have Ethernet connectivity unlike in Model B where we have a RJ45 connector. Consumption for Power in Model A (1.5 W) is less then that of Model B (3.5W).

10.4   Operating system in Raspberry Pi

NOOBS is a way to make setting up a Raspberry Pi for the first time much, much easier. You wont need network access, and you wont need to download any special imaging software. Just head to the https://www.raspberrypi.org/downloads/ page, grab a copy of the NOOBS zip file, and unpack it onto a freshly formatted 4GB (or larger) SD card. When you boot up for the first time, youll see a menu prompting you to install various OSs .

10.5   Why Raspberry Pi ?

If you are a amateur computer and electronics enthusiasts who want to build their own devices this is the best thing you can get at such low cost , its definitely a value for money thing for doing automation. You can pretty much do anything which a low powered computer can do .You can watch a movie, write a document, play basic games, and so onits really up to you.