Programme Educational Objectives(PEOs)

PEO1 (Core Competency) : Graduates will acquire a strong foundation in mathematical, scientific and engineering fundamentals necessary to formulate, solve and analyze Computer Science and Engineering problems.

PEO2 (Professionalism) : Graduates will practice the profession with ethics, integrity and leadership to relate engineering to global perspective issues and social context.

PEO3 (Higher Studies and Entrepreneurship) : Graduates will be prepared for their careers in the software industry or in higher studies leading to research and for applying the spirit of innovation and entrepreneurship in their career and continuing to develop their professional knowledge on a life long basis.

Programme Outcomes(POs)

PO1: Engineering knowledge: Ability to apply the knowledge of mathematics, physical sciences and computer science and engineering specialization to the solution of complex engineering problems.

PO2: Problem analysis: Ability to identify, formulate and analyze complex real life problems in order to provide meaningful solutions by applying knowledge acquired in computer science and engineering.

PO3: Design/development of solutions: Ability to design cost effective software / hardware solutions to meet desired needs of customers/clients.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions in the field of computer science and engineering.

PO5: Modern tool usage: Create, select and apply appropriate techniques, resources and modern computer science and engineering tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Programme Specific Outcomes (PSOs)

PSO1: Software System Design and Development: The ability to apply software development life cycle principles to design and develop the application software that meet the automation needs of society and industry.
PSO2: Computing and Research ability: The ability to employ modern computer languages, environments and platforms in creating innovative career paths in SMAC (Social, Mobile, Analytics and Cloud) technologies.

☢ There is no full automated data science. You need to get your hands dirty ☢

K S R INSTITUTE FOR ENGINEERING AND TECHNOLOGY
K.S.R KALVI NAGAR, TIRUCHENGODE - 637 215
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
KSRIET
TECH ZEST
Fuel Your Mind
Jan- 2018 - Volume V / Issue II

Data Science

K S R Institute for Engineering and Technology

Vision

To become a globalLy recognized Institution in Engineering Education, Research and Entrepreneurship.

Mission

IM1: Accomplish quality education through improved teaching learning process.
IM2: Enrich technical skills with state of the art laboratories and facilities.
IM3: Enhance research and entrepreneurship activities to meet the industrial and societal needs

Department of Computer Science and Engineering

Vision

To produce globally competitive Computer Science Engineers and Entrepreneurs with moral values.

Mission

DM1 (Quality Education) : Provide quality education to enhance problem solving skills, leadership qualities, team spirit and ethical responsibilities.

DM2 (State of art Laboratory) : Enable the students to adapt to the rapidly changing technologies by providing advanced laboratories and facilities.

DM3 (Research and Development) : Promote research based activities in the emerging areas of techno-environment in order to meet industrial and societal needs.

☢ Analytics can be classified into three broad categories descriptive analytics, predictive analytics and prescriptive analytics. ☢

Introduction to Data Science

Data Science is perfect amalgamation of programming, analytical, and business skills that allows extracting meaningful insights from raw and unstructured data.

With enormous amount of facts generating each minute, the requirement to extract the useful insights is a must for the businesses to stand out from the crowd. Data engineers setup the database and data storage in order to facilitate the process of data mining and other processes. Every other organization is running behind profits, but the companies that formulate efficient strategies based on fresh and useful insights always win the game in the long-run.

Data Scientist: "Data Scientist is better at statistics than any software engineer and better at software engineering than any statistician." — Josh Wills, Director of Data Engineering at Slack

Data Analytic Life cycle:

- Understand the Problem
- Collect Enough Data
- Process the Raw Data
- Explore the Data
- Analyze the Data
- Communicate the Results

Advantages of Data Science:

- Data scientist helps the management to come up with better and faster decisions
- It empowers the decision-makers with solid data
- You can anticipate new challenges and opportunities through the power of data
- Spotting trends and capitalizing on it before the competition
- Setting the guidelines for best practices and tried and tested methodologies

What is the difference between an analyst and a data scientist?

"Analyst" is somewhat of an ambiguous job title that can represent many different types of roles (data analyst, marketing analyst, operations analyst, financial analyst, etc). What does this mean in comparison to data scientist?

- Data Scientist: Specialty role with abilities in math, technology, and business acumen. Data scientists work at the raw database level to derive insights and build data product.
- Analyst: This can mean a lot of things. Common thread is that analysts look at data to try to gain insights. Analysts may interact with data at both the database level or the summarized report level.

Let's have a look at the below infographic to see all the domains where Data Science is creating its impression.

S.Niveditha, IV / CSE

☢ **R is one of the most popular programming languages in Data Science.** ☢

Data Science Trends in 2018

Data Science Trend 1 – Artificial Intelligence: The widespread adoption of Artificial Intelligence into all business systems and decision-making applications. According to Gartner, 59 percent of organizations are still building their enterprise AI strategies while the remaining 41 percent of the organizations have already made the plunge.

Data Science Trend 2 – Smart Apps: The next several years will witness a steady rise in AI-driven apps and services. All managed software platforms like the ERP are currently on a race for AI integration in their existing systems for enhanced performance and value addition. This trend includes the use of digital assistants and virtual services

Data Science Trend 3 – Intelligent Things: The Intelligent Things are semi-robotic, smarter versions of regular gadgets and equipments to make our lives easy. They will continue to expand at ever greater rates and in all parts of our lives.

Data Science Trend 4 – Digital Twins: Digital Twins will bring together the connected world of sensors and humans. This technology trend will further the case of mechanized asset management.

"In the context of IoT, digital twins are linked to real-world objects and offer information on the state of the counterparts, respond to changes, improve operations and add value."

Data Science Trend 5 – Edge Computing: Edge Computing "describes a computing topology in which information processing and content collection and delivery are placed closer to the sources of this information." Such technology is directly related to the Internet of Things will expand with far-reaching implication on "sources of information."

Data Science Trend 6 – Intelligent Platforms: The conversational platform pushes machine intelligence up a notch, where human expectations from digital systems will significantly rise. These systems, however, are expected to deliver results based on the event models and APIs they are fed with.

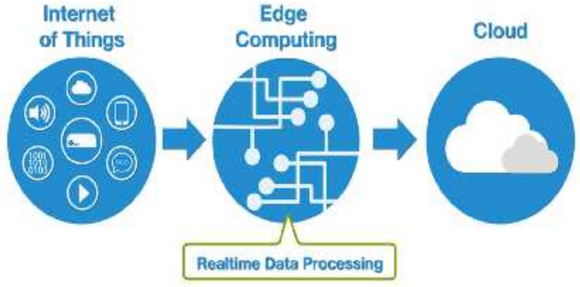
Data Science Trend 7 – Augmented Reality: The Immersive Experience related to augmented reality (AR) and virtual reality (VR) is already changing the world around us. The human-machine interaction will improve as research breakthroughs in AR and VR come about.

Data Science Trend 8 – Blockchain: Blockchain will become a much more important technology for businesses across the globe. Very simply put, Blockchain enables un-trusted parties to engage in transactions. Blockchain holds promise for many industry sectors like the finance, healthcare, and content delivery. This technology is still maturing.

Data Science Trend 9 – Event Driven Techs: Events drive businesses. This trend is expected to bring about some revolutionary changes to joint stakeholders in businesses.

"Some business events or combinations of events constitute business moments — a detected situation that calls for some specific business action. The most consequential business moments are those that have implications for multiple parties, such as separate applications, lines of business or partners."

Data Science Trend 10 – Security: In an ever more complex security environment of digital businesses, a sophisticated tech strategy known as "Continuous Adaptive Risk and Trust Assessment" (CARTA) will enable better decision making with adaptive responses to digital business. The basic premise of CARTA is trust.



S.Thirumurugan
M.Vasanthasenan, IV / CSE

☢ **The bad news? 51% of a data scientist's time is spent collecting, labeling, cleaning and organizing data** ☢

Business analytics (BA)

Business analytics (BA) is the practice of iterative, methodical exploration of an organization's data, with an emphasis on statistical analysis. Business analytics is used by companies committed to data-driven decision-making.

BA is used to gain insights that inform business decisions and can be used to automate and optimize business processes. Data-driven companies treat their data as a corporate asset and leverage it for a competitive advantage. Successful business analytics depends on data quality, skilled analysts who understand the technologies and the business, and an organizational commitment to data-driven decision-making.

Business analytics examples

Business analytics techniques break down into two main areas.

The first is basic business intelligence. This involves examining historical data to get a sense of how a business department, team or staff member performed over a particular time. This is a mature practice that most enterprises are fairly accomplished at using.

The second area of business analytics involves deeper statistical analysis. This may mean doing predictive analytics by applying statistical algorithms to historical data to make a prediction about future performance of a product, service or website design change. Or, it could mean using other advanced analytics techniques, like cluster analysis, to group customers based on similarities across several data points.

Specific types of business analytics include:

Descriptive analytics, which tracks key performance indicators to understand the present state of a business;

Predictive analytics, which analyzes trend data to assess the likelihood of future outcomes; and Prescriptive analytics, which uses past performance to generate recommendations about how to handle similar situations in the future.

S.Gowri, II / CSE

Data Science in Airline Route Planning

Airline Industry across the world is known to bear heavy losses. Except a few airline service providers, companies are struggling to maintain their occupancy ratio and operating profits. With high rise in air fuel prices and need to offer heavy discounts to customers has further made the situation worse. It wasn't for long when airlines companies started using data science to identify the strategic areas of improvements. Now using data science, the airline companies can:

1. Predict flight delay
2. Decide which class of airplanes to buy
3. Whether to directly land at the destination, or take a halt in between (For example: A flight can have a direct route from New Delhi to New York. Alternatively, it can also choose to halt in any country.)
4. Effectively drive customer loyalty programs

M.Sangeetha, III / CSE

Data Science in Fraud And Risk Detection

One of the first applications of data science originated from Finance discipline. Companies were fed up of bad debts and losses every year. However, they had a lot of data which use to get collected during the initial paper work while sanctioning loans. They decided to bring in data science practices in order to rescue them out of losses. Over the years, banking companies learned to divide and conquer data via customer profiling, past expenditures and other essential variables to analyze the probabilities of risk and default. Moreover, it also helped them to push their banking products based on customer's purchasing power.

S.Subhasree, II / CSE

Each minute up to 300 hours of video are uploaded to YouTube!

Data Science in Finance

Automating risk management

Risk management is an enormously important area for financial institutions, responsible for company's security, trustworthiness, and strategic decisions. The approaches to handling risk management have changed significantly over the past years, transforming the nature of finance sector. As never before, machine learning models today define the vectors of business development.

There are many origins from which risks can come, such as competitors, investors, regulators, or company's customers. Also, risks can differ in importance and potential losses. Therefore, the main steps are identifying, prioritizing, and monitoring risks, which are the perfect tasks for machine learning. With training on the huge amount of customer data, financial lending, and insurance results, algorithms can not only increase the risk scoring models but also enhance cost efficiency and sustainability.

Among the most important applications of data science and artificial intelligence (AI) in risk management is identifying the creditworthiness of potential customers. To establish the appropriate credit amount for a particular customer, companies use machine learning algorithms that can analyze past spending behavior and patterns. This approach is also useful while working with new customers or the ones with a brief credit history.

Although digitalization and automatization of risk management processes in finance are in the early stages, the potential is extremely huge. Financial institutions still need to prepare for this change by automating core financial processes, improving analytical skills of the finance team, and making strategic technology investments. But as soon as the company starts to move in this direction, the profit will not make itself wait.

R.manjula, II / CSE

Managing customer data

For financial firms, data is the most important resource. Therefore, efficient data management is a key to business success. Today, there is a massive volume of financial data diversity in structure and volume: from social media activity and mobile interactions to market data and transaction details. Financial specialists often have to work with semi-structured or unstructured data and there is a big challenge to process it manually.

However, it's obvious for most companies that integrating machine learning techniques to managing process is simply a necessity to extract real intelligence from data. AI tools, in particular, natural language processing, data mining, and text analytics, help to transform data into information contributing in smarter data governance and better business solutions, and as a result - increased profitability. For instance, machine learning algorithms can analyze the influence of some specific financial trends and market developments by learning from customers financial historical data. Finally, these techniques can be used to generate automated reports.

A.Soundharya, II / CSE

Predictive analytics

Analytics is now at the core of financial services. Special attention deserves predictive analytics that reveals patterns in the data that foresee the future event that can be acted upon now. Through understanding social media, news trends, and other data sources these sophisticated analytics conquered the main applications such as predicting prices and customers lifetime value, future life events, anticipated churn, and the stock market moves. Most importantly such techniques can help answer the complicated question - how best to intervene.

R.S.Vasumitha, II / CSE

Facebook users send on average 31.25 million messages and view 2.77 million videos every minute.

R

R is a programming language and software environment for statistical analysis, graphics representation and reporting. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team.

The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions. R allows integration with the procedures written in the C, C++, .Net, Python or FORTRAN languages for efficiency.

R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems like Linux, Windows and Mac.

R is free software distributed under a GNU-style copy left, and an official part of the GNU project called GNU S.

Features of R

R is a well-developed, simple and effective programming language which includes conditionals, loops, user defined recursive functions and input and output facilities.

R has an effective data handling and storage facility,

R provides a suite of operators for calculations on arrays, lists, vectors and matrices.

R provides a large, coherent and integrated collection of tools for data analysis.

R provides graphical facilities for data analysis and display either directly at the computer or printing at the papers.

Applications of R Programming in Real World

Data Science

Harvard Business Review named data scientist the "sexiest job of the 21st century". Glassdoor named it the "best job of the year" for 2016. With the advent of IoT devices creating terabytes and terabytes of data that can be used to make better decisions, data science is a field that has no other way to go but up.Simply explained, a data scientist is a statistician with an extra asset: computer programming skills. Programming languages like R give a data scientist superpowers that allow them to collect data in realtime, perform statistical and predictive analysis, create visualizations and communicate actionable results to stakeholders.Most courses on data science include R in their curriculum because it is the data scientist's favourite tool.

Statistical computing

R is the most popular programming language among statisticians. In fact, it was initially built by statisticians for statisticians. It has a rich package repository with more than 9100 packages with every statistical function you can imagine.R's expressive syntax allows researchers – even those from non computer science backgrounds to quickly import, clean and analyze data from various data sources.R also has charting capabilities, which means you can plot your data and create interesting visualizations from any dataset.

Machine Learning

R has found a lot of use in predictive analytics and machine learning. It has various package for common ML tasks like linear and non-linear regression, decision trees, linear and non-linear classification and many more.Everyone from machine learning enthusiasts to researchers use R to implement machine learning algorithms in fields like finance, genetics research, retail, marketing and health care.

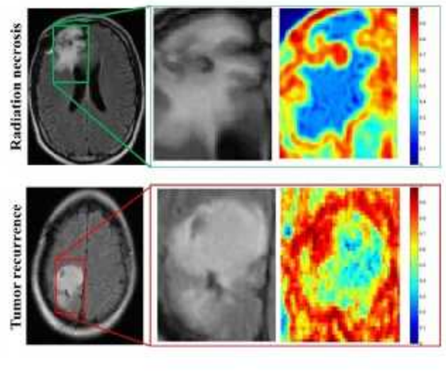
P.Keerthana , G.Chithradevi, IV / CSE

Data Science in Healthcare

Medical image analysis

The healthcare sector receives great benefits from the data science application in medical imaging. There is a lot of research in this area, and one of the major studies is Big Data Analytics in Healthcare, published in BioMed Research International. According to the study, popular imaging techniques include magnetic resonance imaging (MRI), X-ray, computed tomography, mammography, and so on. Numerous methods are used to tackle the difference in modality, resolution, and dimension of these images. Many more are being developed to improve the image quality, extract data from images more efficiently, and provide the most accurate interpretation. The deep-learning based algorithms increase the diagnostic accuracy by learning from the previous examples and then suggest better treatment solutions.

The most popular image-processing techniques focus on enhancement, segmentation, and denoising that allows deep analysis of organ anatomy, and detection of diverse disease conditions.



The most promising applications aim to detect tumors, artery stenosis, organ delineation, etc. Different methods and frameworks contribute to medical imaging in various aspects. Hadoop, a popular analytical framework, employs MapReduce to find the optimal parameters for tasks like lung texture classification. It applies machine learning methods, support vector machines (SVM), content-based medical image indexing, and wavelet analysis for solid texture classification.

Other examples include iDASH (integrating data for analysis, anonymization, and sharing) used for biomedical computing, HAMSTER/MPI GraphLabfor processing large images, and more.

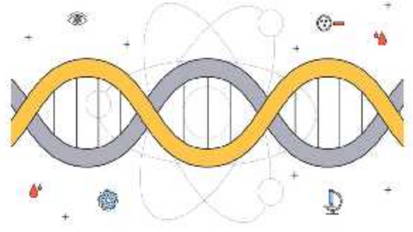
A. Pavithra, III / CSE

Genetics and Genomics

The research in genetics and genomics enables an advanced level of treatment personalization. The goal is to understand the impact of the DNA on our health and find individual biological connections between genetics, diseases, and drug response. Data science techniques allow integration of different kinds of data with genomic data in the disease research, which provides a deeper understanding of genetic issues in reactions to particular drugs and diseases. Let us review the most popular techniques and frameworks.

MapReduce allows reading genetic sequences mapping and shortens the time for efficient data processing. SQL contributes to retrieving genomic data, BAM file manipulations, and computation. The Deep Genomics made a remarkable impact on predicting the molecular effects of genetic variation essential to DNA interpretation. Their database has enabled the scientists to understand how genetic variations can impact a genetic code.

Despite the significant progress in developing the DNA sequencing technologies in the recent years, a lotis still left to explore, and the perspectives look encouraging. Many challenges remaindue to the continuous interactions between genes and the external variables. As soon as we acquire a reliable personal genome data, we will achieve a deeper understanding of the human DNA. The advanced genetic risk prediction will be a major step towards more individual care.



A. Janani, III / CSE