CodeForce Course on Data Science - Summer 2020

Data science is an exciting discipline that allows you to turn raw data into understanding, insight, and knowledge. The course on Data Science is divided into 5 core modules which will bring the students up to speed with the latest developments in the industry and help them break into the sector. This course is designed to help beginners get an understanding of what data science is, intermediate to enhance their core skills, and advanced to implement complex data pipelines.

The course is structured to be a very fast-paced deep dive into all the relevant practical topics in the Data Science sector. There will be classroom teaching followed by assignments and projects that the students will be expected to go through. Each class will start with a discussion on the pre-class readings, will cover several theoretical concepts during the lecture along with code implementations of all the concepts, and will be followed by extensive assignments that the students are expected to submit before the start of the next class. The list of references, textbooks, and other materials will be provided to the students for each class.

Instructor: Suraj Peri, email: suraj.peri93@gmail.com
Schedule: Mon to Fri, 11 AM to 1 PM EST

Suraj is the co-founder of Gradvine Advisors Private Limited. He is an alumnus of Carnegie Mellon University where he pursued his Masters with a specialization in Data Science and Artificial Intelligence. Post completing his Masters, Suraj worked at the University of California, Berkeley, as an Energy Policy analyst working at the intersection of data science and energy policy and consulting for senior Senator's staff and the Governor's office in the state of California. While at Berkeley, Suraj went on to specialise and take courses in core data science and training from the University of California, Berkeley, and a course on Advanced Statistical Learning from Stanford University. He went onto work at Itron, an energy consulting firm, leading the Data Science team and working on advanced machine learning techniques as applied to the energy industry. Suraj has since moved back to India to take over as Chief Operating Officer at Gradvine and runs workshops to teach high school and college students the art and science of dealing with data. He has trained thousands of students in core Data Sciences and his students have gone on to places like Google, Facebook, Oracle, Uber, and hundreds of other companies.

Important pep talk from the instructor -

I promise you can succeed in this class.

Learning Data Science can be difficult at first—it's like learning a new language, just like Spanish, French, or Chinese. Hadley Wickham—the chief data scientist at RStudio and the author of some amazing R packages you'll be using like ggplot2—made this wise observation:

It's easy when you start out programming to get really frustrated and think, "Oh it's me, I'm really stupid," or, "I'm not made out to program." But, that is absolutely not the case. Everyone gets frustrated. I still get frustrated occasionally when writing R code. It's just a natural part of programming. So, it happens to everyone and gets less and less over time. Don't blame yourself. Just take a break, do something fun, and then come back and try again later.

Even experienced programmers find themselves bashing their heads against seemingly intractable errors. If you're finding yourself taking way too long hitting your head against a wall and not understanding, take a break, talk to classmates, e-mail me, etc.
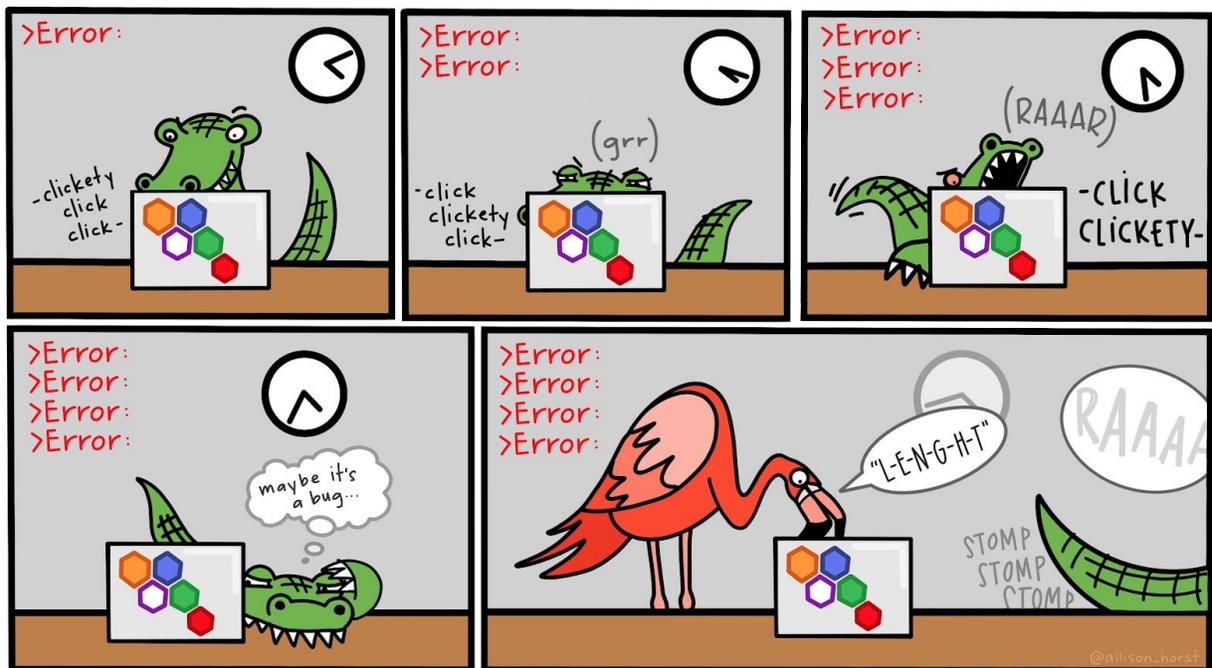
Happy Learning!

Course Content -

The class will be divided into the following 5 core modules. All the topics to be covered in the individual modules will be provided to the students prior to the start of the module itself.

CodeForce DS01 - Introduction to Data Analytics

This course teaches the fundamental ideas to clean, manipulate, process, and analyze data. The students will work on data analysis problems arising in various data-intensive applications. The course involves many in-class coding exercises where the students are expected to work on several case studies. Through these exercises, the course shall also serve as an introduction to data analytics and modern scientific computing with Python and R programming languages, along with data storage and extraction from databases.

CodeForce DS02 - Applied Statistics from a Practitioner's Point of View

This course is an applied statistics course with an emphasis on data analysis. In this course, we will study several statistical modeling techniques and discuss real-life problems over which we'll have a chance to apply statistical tools to learn from data. We will be covering some of the fundamental statistical methods like linear regression, principal component analysis, cross-validation, and p-values. The lectures are designed to help the participants apply these techniques on large sets of data using a statistical programming language such as R or Python.

CodeForce DS03 - Exploration and Data Visualisation

Data rarely speaks for itself. On their own, the facts contained in raw data are difficult to understand, and in the absence of beauty and order, it is impossible to understand the truth that the data shows. In this module, you'll learn how to use industry-standard graphic and data design techniques to create beautiful, understandable visualizations and uncover the truth in data. By the end of this course, you will become (1) literate in data and graphic design principles, and (2) an ethical data communicator, by producing beautiful, powerful, and clear visualizations of your own data.

CodeForce DS04 - Machine Learning

In this course, we will cover the fundamental aspects of Machine Learning. We will start with the fundamentals of machine learning, including different learning paradigms, regression and classification problems, evaluation methods, generalization, and overfitting. We will then cover some of the fundamental machine learning techniques such as decision trees, Bayesian approaches, Naive Bayes classifier, and logistic regression, k-Nearest neighbor, and online learning algorithms. Besides understanding the basic theory behind the techniques, students are expected to apply them using both Python and R. Machine learning aims to develop computer programs that improve their performance through experience by capturing relevant abstractions of past training input. This course will cover fundamental approaches in machine learning (k-NN, decision trees, Bayesian approaches, neural networks, support vector machines, ensemble learning), as well as getting a solid foundation on theoretical concepts such as overfitting, the curse of dimensionality, bias-variance dilemma.

CodeForce DS05 - Deep Learning

Get a crash course on what there is to learn and how to go about learning more. Deep Learning module presents a simplified explanation of some of the hottest topics in data science today:

- What is Deep Learning?
- What are convolutional neural networks?
- Why is deep learning so powerful and what can it be used for?
- Be part of a rapidly growing field in data science; there's no better time than now to get started with neural networks.