tutorialspoint

SIMPLY EASY LEARNING

www.tutorialspoint.com

https://www.facebook.com/tutorialspointindia

https://twitter.com/tutorialspoint

About the Tutorial

Agile is a software development methodology that helps in building software through incremental sessions using short iterations of 1 to 4 weeks so that the development is aligned with the changing business needs. Agile Data science comprises of a combination of agile methodology and data science. In this tutorial, we have used appropriate examples to help you understand agile development and data science in a general and quick way.

Audience

This tutorial has been prepared for developers and project managers to help them understand the basics of agile principles and its implementation. After completing this tutorial, you will find yourself at a moderate level of expertise, from where you can advance further with implementation of data science and agile methodology.

Prerequisites

It is important to have basic knowledge of data science modules and software development concepts such as software requirements, coding along with testing.

Copyright & Disclaimer

Table of Contents

1. Agile Data Science – Introduction

Agile data science is an approach of using data science with agile methodology for web application development. It focusses on the output of the data science process suitable for effecting change for an organization. Data science includes building applications that describe research process with analysis, interactive visualization and now applied machine learning as well.

The major goal of agile data science is to -

document and guide explanatory data analysis to discover and follow the critical path to a compelling product.

Agile data science is organized with the following set of principles:

Continuous Iteration

This process involves continuous iteration with creation tables, charts, reports and predictions. Building predictive models will require many iterations of feature engineering with extraction and production of insight.

Intermediate Output

This is the track list of outputs generated. It is even said that failed experiments also have output. Tracking output of every iteration will help creating better output in the next iteration.

Prototype Experiments

Prototype experiments involve assigning tasks and generating output as per the experiments. In a given task, we must iterate to achieve insight and these iterations can be best explained as experiments.

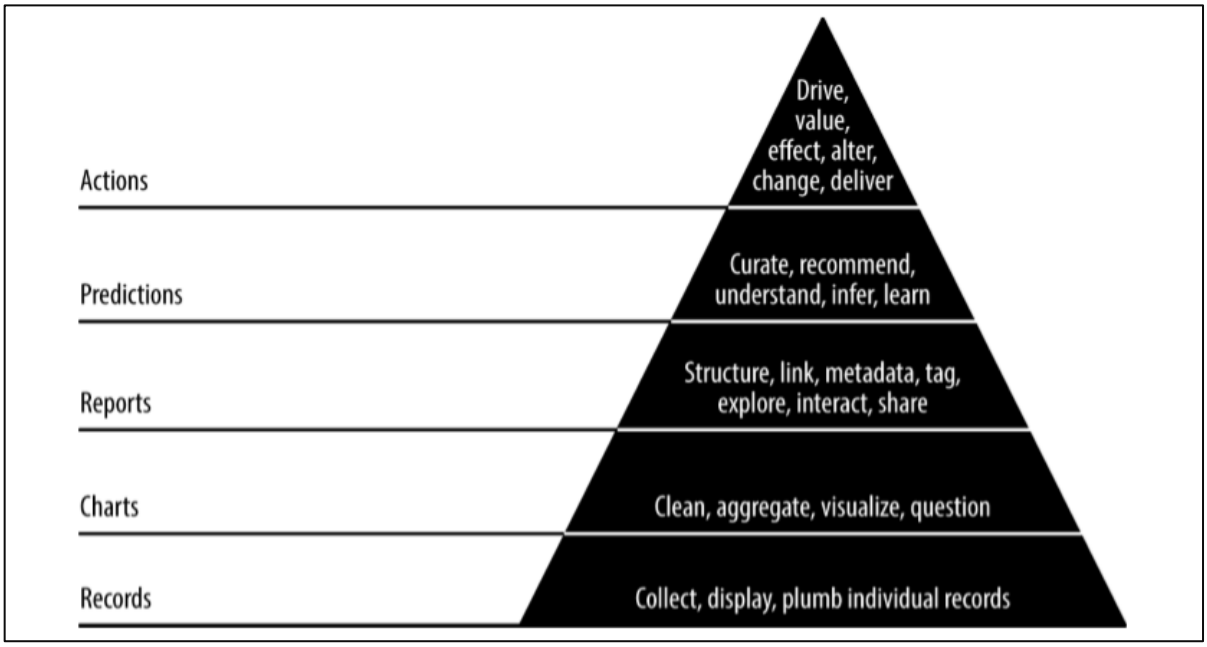
Integration of data

The software development life cycle includes different phases with data essential for –

- customers
- developers, and
- the business

The integration of data paves way for better prospects and outputs.

Pyramid data value



The above pyramid value described the layers needed for "Agile data science" development. It starts with a collection of records based on the requirements and plumbing individual records. The charts are created after cleaning and aggregation of data. The aggregated data can be used for data visualization. Reports are generated with proper structure, metadata and tags of data. The second layer of pyramid from the top includes prediction analysis. The prediction layer is where more value is created but helps in creating good predictions that focus on feature engineering.

The topmost layer involves actions where the value of data is driven effectively. The best illustration of this implementation is "Artificial Intelligence".

2. Agile Data Science – Methodology Concepts

In this chapter, we will focus on the concepts of software development life cycle called "agile". The Agile software development methodology helps in building a software through increment sessions in short iterations of 1 to 4 weeks so the development is aligned with changing business requirements.

There are 12 principles that describe the Agile methodology in detail:

Satisfaction of customers

The highest priority is given to customers focusing on the requirements through early and continuous delivery of valuable software.

Welcoming new changes

Changes are acceptable during software development. Agile processes is designed to work in order to match the customer's competitive advantage.

Delivery

Delivery of a working software is given to clients within a span of one to four weeks.

Collaboration

Business analysts, quality analysts and developers must work together during the entire life cycle of project.

Motivation

Projects should be designed with a clan of motivated individuals. It provides an environment to support individual team members.

Personal conversation

Face-to-face conversation is the most efficient and effective method of sending information to and within a development team.

Measuring progress

Measuring progress is the key that helps in defining the progress of project and software development.

Maintaining constant pace

Agile process focusses on sustainable development. The business, the developers and the users should be able to maintain a constant pace with the project.

Monitoring

It is mandatory to maintain regular attention to technical excellence and good design to enhance the agile functionality.

Simplicity

Agile process keeps everything simple and uses simple terms to measure the work that is not completed.

Self-organized terms

An agile team should be self-organized and should be independent with the best architecture; requirements and designs emerge from self-organized teams.

Review the work

It is important to review the work at regular intervals so that the team can reflect on how the work is progressing. Reviewing the module on a timely basis will improve performance.

Daily Stand-up

Daily stand-up refers to the daily status meeting among the team members. It provides updates related to the software development. It also refers to addressing obstacles of project development.

Daily stand-up is a mandatory practice, no matter how an agile team is established regardless of its office location.

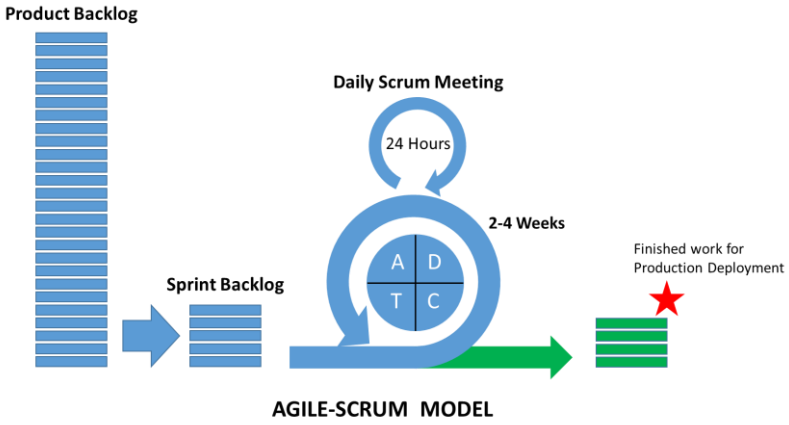
The list of features of a daily stand-up are as follows:

- The duration of daily stand-up meet should be roughly 15 minutes. It should not extend for a longer duration.
- Stand-up should include discussions on status update.
- Participants of this meeting usually stand with the intention to end up meeting quickly.

User Story

A story is usually a requirement, which is formulated in few sentences in simple language and it should be completed within an iteration. A user story should include the following characteristics:

- All the related code should have related check-ins.
- The unit test cases for the specified iteration.
- All the acceptance test cases should be defined.
- Acceptance from product owner while defining the story.



What is Scrum?

Scrum can be considered as a subset of agile methodology. It is a lightweight process and includes the following features:

- It is a process framework, which includes a set of practices that need to be followed in consistent order. The best illustration of Scrum is following iterations or sprints.
- It is a "lightweight" process meaning that the process is kept as small as possible, to maximize the productive output in given duration specified.

Scrum process is known for its distinguishing process in comparison with other methodologies of traditional agile approach. It is divided into the following three categories:

- Roles
- Artifacts
- Time Boxes

Roles define the team members and their roles included throughout the process. The Scrum Team consists of the following three roles:

- Scrum Master
- Product Owner
- Team

The Scrum artifacts provide key information that each member should be aware of. The information includes details of product, activities planned, and activities completed. The artefacts defined in Scrum framework are as follows:

- Product backlog
- Sprint backlog
- Burn down chart
- Increment

Time boxes are the user stories which are planned for each iteration. These user stories help in describing the product features which form part of the Scrum artefacts. The product backlog is a list of user stories. These user stories are prioritized and forwarded to the user meetings to decide which one should be taken up.

Why Scrum Master?

Scrum Master interacts with every member of the team. Let us now see the interaction of the Scrum Master with other teams and resources.

Product Owner

The Scrum Master interacts the product owner in following ways:

- Finding techniques to achieve effective product backlog of user stories and managing them.
- Helping team to understand the needs of clear and concise product backlog items.
- Product planning with specific environment.
- Ensuring that product owner knows how to increase the value of product.
- Facilitating Scrum events as and when required.

Scrum Team

The Scrum Master interacts with the team in several ways:

- Coaching the organization in its Scrum adoption.
- Planning Scrum implementations to the specific organization.
- Helping employees and stakeholders to understand the requirement and phases of product development.
- Working with Scrum Masters of other teams to increase effectiveness of the application of Scrum of the specified team.

Organization

The Scrum Master interacts with organization in several ways. A few are mentioned below:

- Coaching and scrum team interacts with self-organization and includes a feature of cross functionality.
- Coaching the organization and teams in such areas where Scrum is not fully adopted yet or not accepted.

Benefits of Scrum

Scrum helps customers, team members and stakeholders collaborate. It includes time-boxed approach and continuous feedback from the product owner ensuring that the product is in working condition. Scrum provides benefits to different roles of the project.

Customer

The sprints or iterations are considered for shorter duration and user stories are designed as per priority and are taken up at sprint planning. It ensures that every sprint delivery, customer requirements are fulfilled. If not, the requirements are noted and are planned and taken for sprint.

Organization

Organization with the help of Scrum and Scrum masters can focus on the efforts required for development of user stories thus reducing work overload and avoiding rework if any. This also helps in maintaining increased efficiency of development team and customer satisfaction. This approach also helps in increasing the potential of the market.

Product Managers

The main responsibility of the product managers is to ensure that the quality of product is maintained. With the help of Scrum Masters, it becomes easy to facilitate work, gather quick responses and absorb changes if any. Product managers also verify that the designed product is aligned as per the customer requirements in every sprint.

Development Team

With time-boxed nature and keeping sprints for a smaller duration of time, development team becomes enthusiastic to see that the work is reflected and delivered properly. The working product increments each level after every iteration or rather we can call them as "sprint". The user stories which are designed for every sprint become customer priority adding up more value to the iteration.

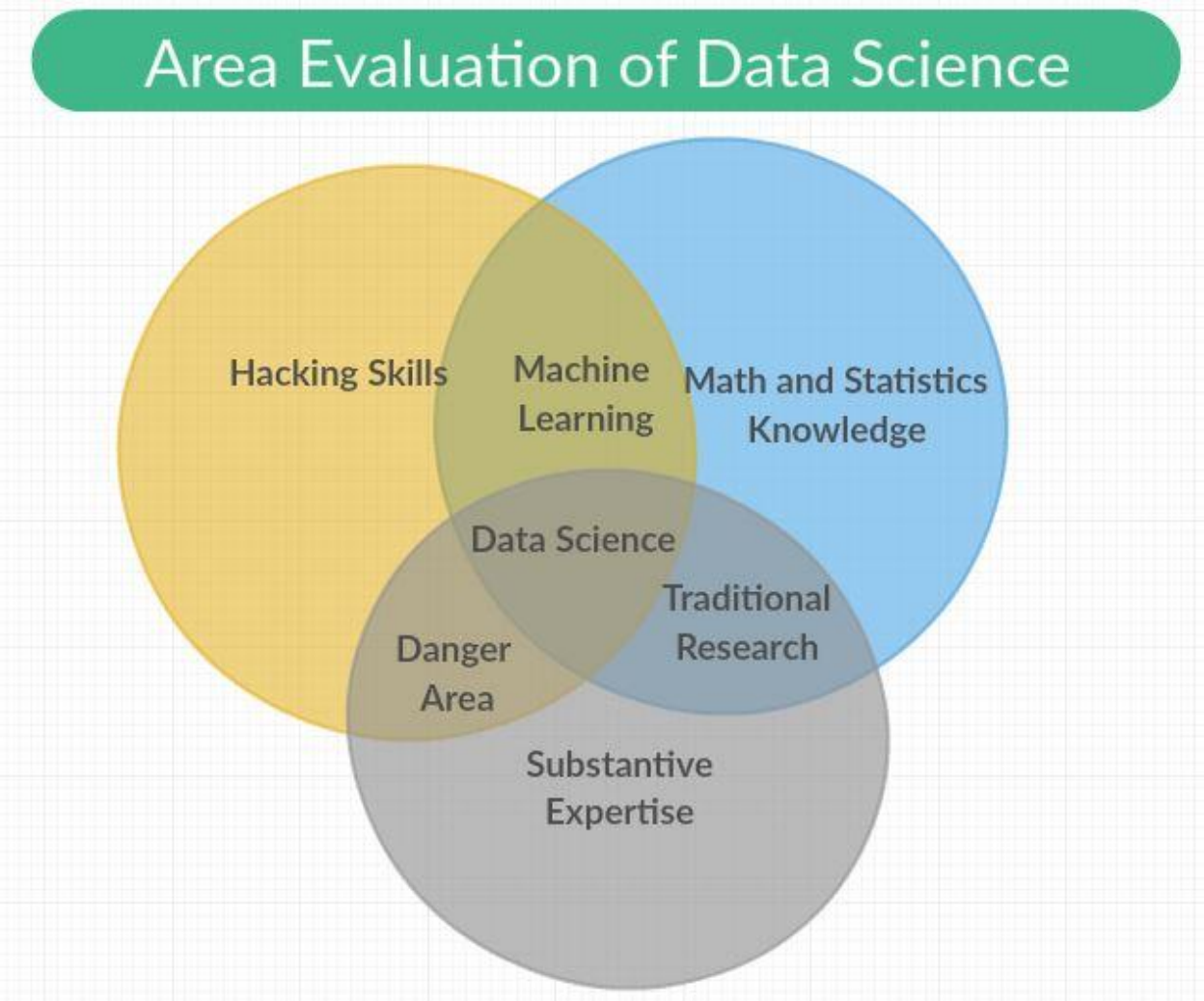
Conclusion

Scrum is an efficient framework within which you can develop software in teamwork. It is completely designed on agile principles. ScrumMaster is there to help and co-operate the team of Scrum in every possible way. He acts like a personal trainer who helps you stick with designed plan and perform all the activities as per the plan. The authority of ScrumMaster should never extend beyond the process. He/she should be potentially capable to manage every situation.

3. Agile Data Science – Data Science Process

In this chapter, we will understand the data science process and terminologies required to understand the process.

"Data science is the blend of data interface, algorithm development and technology in order to solve analytical complex problems".



Data science is an interdisciplinary field encompassing scientific methods, processes and systems with categories included in it as Machine learning, math and statistics knowledge with traditional research. It also includes a combination of hacking skills with substantive expertise. Data science draws principles from mathematics, statistics, information science, and computer science, data mining and predictive analysis.

The different roles that form part of the data science team are mentioned below:

Customers

Customers are the people who use the product. Their interest determines the success of project and their feedback is very valuable in data science.

Business Development

This team of data science signs in early customers, either firsthand or through creation of landing pages and promotions. Business development team delivers the value of product.

Product Managers

Product managers take in the importance to create best product, which is valuable in market.

Interaction designers

They focus on design interactions around data models so that users find appropriate value.

Data scientists

Data scientists explore and transform the data in new ways to create and publish new features. These scientists also combine data from diverse sources to create a new value. They play an important role in creating visualizations with researchers, engineers and web developers.

Researchers

As the name specifies researchers are involved in research activities. They solve complicated problems, which data scientists cannot do. These problems involve intense focus and time of machine learning and statistics module.

Adapting to Change

All the team members of data science are required to adapt to new changes and work on the basis of requirements. Several changes should be made for adopting agile methodology with data science, which are mentioned as follows:

- Choosing generalists over specialists.
- Preference of small teams over large teams.
- Using high-level tools and platforms.
- Continuous and iterative sharing of intermediate work.

Note:

In the Agile data science team, a small team of generalists uses high-level tools that are scalable and refine data through iterations into increasingly higher states of value.

Consider the following examples related to the work of data science team members:

- Designers deliver CSS.
- Web developers build entire applications, understand the user experience, and interface design.
- Data scientists should work on both research and building web services including web applications.
- Researchers work in code base, which shows results explaining intermediate results.
- Product managers try identifying and understanding the flaws in all the related areas.

4. Agile Data Science – Agile Tools and Installation

In this chapter, we will learn about the different Agile tools and their installation. The development stack of agile methodology includes the following set of components:

Events

An event is an occurrence that happens or is logged along with its features and timestamps.

An event can come in many forms like servers, sensors, financial transactions or actions, which our users take in our application. In this complete tutorial, we will use JSON files that will facilitate data exchange among different tools and languages.

Collectors

Collectors are event aggregators. They collect events in a systematic manner to store and aggregate bulky data queuing them for action by real time workers.

Distributed document

These documents include multinode (multiple nodes) which stores document in a specific format. We will focus on MongoDB in this tutorial.

Web application server

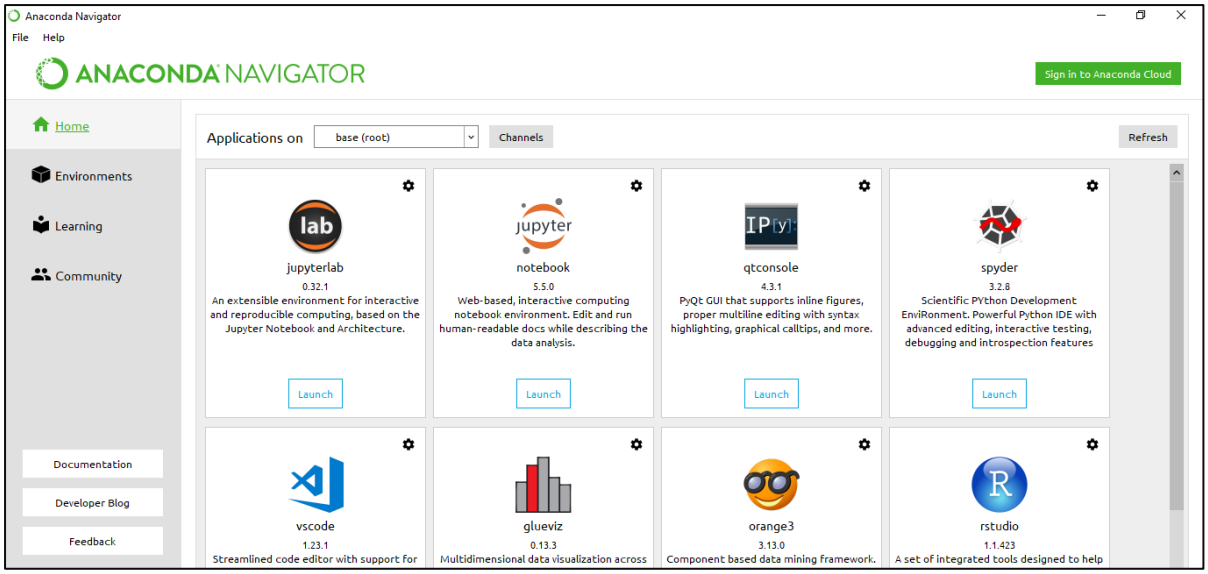
Web application server enables data as JSON through client through visualization, with minimal overhead. It means web application server helps to test and deploy the projects created with agile methodology.

Modern Browser

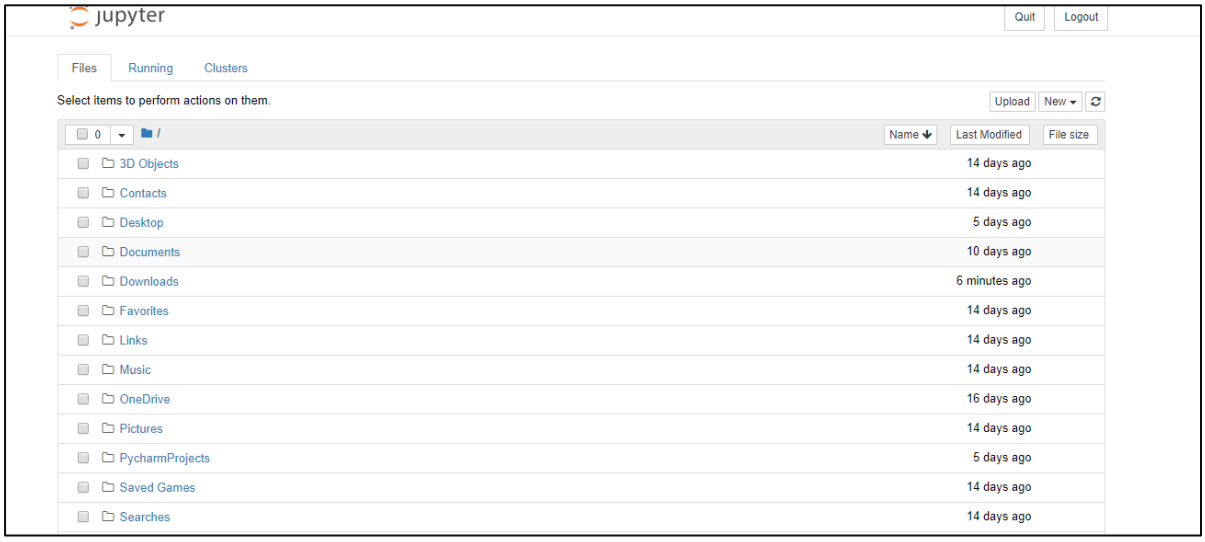
It enables modern browser or application to present data as an interactive tool for our users.

Local Environmental Setup

For managing data sets, we will focus on the Anaconda framework of python that includes tools for managing excel, csv and many more files. The dashboard of Anaconda framework once installed is as shown below. It is also called the "Anaconda Navigator":



The navigator includes the "Jupyter framework" which is a notebook system that helps to manage datasets. Once you launch the framework, it will be hosted in the browser as mentioned below:



End of ebook preview

If you liked what you saw…

Buy it from our store @ **https://store.tutorialspoint.com**