

PYTHON DATA SCIENCE

LEARN THE ETHIC OF CODING IN A
DAY BY TAKING MY CLASSES



TIM WIRED

Python Data Science

*Learn the Ethics of Coding in a Day
by Taking My Classes*

TIM WIRED

Ó Copyright 2019 by Tim Wired - All rights reserved.

This eBook is provided with the sole purpose of providing relevant information on a specific topic for which every reasonable effort has been made to ensure that it is both accurate and reasonable. Nevertheless, by purchasing this eBook you consent to the fact that the author, as well as the publisher, are in no way experts on the topics contained herein, regardless of any claims as such that may be made within. As such, any suggestions or recommendations that are made within are done so purely for entertainment value. It is recommended that you always consult a professional prior to undertaking any of the advice or techniques discussed within.

This is a legally binding declaration that is considered both valid and fair by both the Committee of Publishers Association and the

American Bar Association and should be considered as legally binding within the United States.

The reproduction, transmission, and duplication of any of the content found herein, including any specific or extended information will be done as an illegal act regardless of the end form the information ultimately takes. This includes copied versions of the work both physical, digital and audio unless express consent of the Publisher is provided beforehand. Any additional rights reserved.

Furthermore, the information that can be found within the pages described forthwith shall be considered both accurate and truthful when it comes to the recounting of facts. As such, any use, correct or incorrect, of the provided information will render the Publisher free of responsibility as to the actions taken outside of their direct purview. Regardless, there are zero scenarios where the original author or the Publisher can be deemed liable in any fashion for any damages or hardships that may result from any of the information discussed herein.

Additionally, the information in the following pages is intended only for informational purposes and should thus be thought of as universal. As befitting its nature, it is presented without assurance regarding its prolonged validity or interim quality. Trademarks that are mentioned are done without written consent and can in no way be considered an endorsement from the trademark holder.

Table of Contents

Introduction

Chapter 1: The Basics of Data Science

Why Does My Business Need Data Science?
A Look at Data Science
Data Science and Business Intelligence
The Lifecycle of Data Science

Chapter 2: How Does the Python Language Fit In With Data Science?

Chapter 3: The Best Python Libraries to Help with Data Science

NumPy
SciPy
Pandas
Matplotlib
Scikit-Learn
Theano
TensorFlow
Keras

Chapter 4: Gathering Your Data

Know Your Biggest Business Problem
Places to Look for the Data
Where to Store the Data?

Chapter 5: Organizing and Cleaning the Data

What is Data Preparation?
Why Is Data Preparation So Important?
Steps Involved for Data Preparation

Chapter 6: A Look at Data Mining

The Process of Data Mining
A Look at Data Warehousing

Chapter 7: Adding Machine Learning to the Mix

Why Should I Use Machine Learning?
Supervised Machine Learning
Unsupervised Machine Learning
Reinforcement Machine Learning

Chapter 8: Completing the Data Analysis

What is Data Analysis?

Steps in a Data Analysis?

Chapter 9: The Importance of Data Visualizations to Finish the Process

Chapter 10: How Data Science Can Take Businesses into the Future

Helps You Learn More About Your Customers

Cut Out Fraud and Other Issues

Learn How to Cut Out Waste

How to Handle Your Competition

Make Better Business Decisions

Conclusion

Description

Introduction

Congratulations on purchasing *Python Data Science* and thank you for doing so.

The following chapters are going to give us a more in-depth look at what we need in order to get things done when it is time to handle any project with data science. Many companies are jumping on board when it is time to work with data science and all that it can provide to us in terms of beating out the competition, learning more about the competition, and so much more. There isn't a single industry out there that is not able to benefit from working with a data science look at themselves, and this guidebook is going to spend some time exploring this in more depth to help us get started.

At the beginning of this guidebook, we are going to spend some time exploring more about data science and what we are able to do with it to help us reach our goals. We will also take a look at some of the benefits of working with this process, and how it can all come together when we add in the Python language to the mix as well.

From there it is time to spend more of our energy on some of the different steps that we need to use in order to make data science success for our needs. For example, we will spend some time taking a look at some of the libraries that are available through Python that can really get the work done for you, look at some of the steps that come with organizing and cleaning your data before using it, and even why it is so important to gather the right kinds and amounts of data to get some of the results that you would like.

There are just so many parts that come into the data science project, and simply set it all up so that the information is ready to go after your analysis. If the data is not gathered, and you are not able to clean up and handle some of the outliers that are there, then you will find that it is really hard to go through and actually get the accurate results that you need from a data science project.

Once we have all of that done, it is time to get into the importance of data analysis. Not only will we take a look at some of the steps that you need to use in order to get started with this process of data analysis, but we will look at how machine learning is able to fit into this process and help us to get so much done in the process. Machine learning, along with the variety of methods and algorithms that go with it, will be so important when it is time to work on any data science project, and will help us to finally take all of that data that we have been working with so far, and expand it so that we are able to really get some good results and know how to use it.

This is not the end of our exploration when we go through this process though. We are also going to spend some time taking a look at data science from the view of why visuals are so important to understanding the data that we have and some of the best visuals that we are able to use to get the results that we want. We can also end this with a look at some of the benefits of working with this process and how data science, when it is used properly, is going to be able to help us to finally get some of the results that we want in the present, and in the future.

There are so many businesses that are going to benefit when it comes to working with data analysis and seeing this work for some of their needs. It is always a good idea to learn a bit more about this process and how you are able to really use it to learn more about your customers, learn how to beat out your competition, and so much more. But we have to learn that there are a lot of steps that come into play, and we can't just skip right to the predictions and insights, no matter how much we would like to do this.

Data science is going to be a process that has a lot of different parts that come together with each other, and when we learn how to make these work, we are going to see some amazing results in the process as well. It is really as simple as all of that! When you follow the steps of gathering the right data, organizing and cleaning your data, data mining, a data analysis, and even adding in some visuals of the data, in the end, you will find that there is so much more than you are able to do with this process overall.

When you are ready to learn about data science and how you are able to benefit from this process you have to go through a number of steps to make this happen. It may seem like a lot of work, but the goal of this guidebook is to help you get started and to ensure that you are ready to take on some of the work that you need along the way. When your business is ready to use this and see some of the great benefits that come with it, make sure to read through this guidebook to help you get started.

There are plenty of books on this subject on the market, thanks again for choosing this one! Every effort was made to ensure it is full of as much useful information as possible, please enjoy it!

Chapter 1: The Basics of Data Science



Looking at the world around us, we have entered the era of big data, the need for its storage is also going to grow. It was one of the biggest challenges that a lot of industries were facing in the past, at least until 2010. The main focus of that time was a bit different at that time compared to what we are doing now, which is to build up a framework and find other solutions that would help to store the data that we have.

Today, there are a lot of options out there that we are able to use when it is time to store the data that we need. And because of this, some of the focus on data has switched over to how we are going to process this data. Data science is the key that we need to focus on when it is time to work with this. All of the ideas that you see in our favorite sci-fi movies can actually turn into reality when we are able to work with the process of data science in the proper manner. Often, we will find that data science is going to be one of the future parts of

artificial intelligence. This is part of why data science is so important for us to understand and we need to see how it is able to add in some more value to our business.

With some of this in mind, we need to take a look at what data science is all about and how we are able to make this work for some of our needs. There are a lot of parts that are going to come into play when we are working with data science and learning more about how it works is going to make all of the difference.

Why Does My Business Need Data Science?



In a traditional sense, the data that we had was going to be small and structured in size. This is going to make it easier to work with and allows us to get all of this done with some of the simple business intelligence tools that we need. But unlike the data that we are used to seeing in a traditional system, most of the data that we are going to find today is going to be either semi-structured or unstructured.

This makes it a bit more complicated to work with and can cause us to need to accomplish a bit more work in order to see it happen as well.

The data that we are going to work with today, this data is going to come to us in a lot of different sources like text files, multimedia forms, financial logs, and sensors to name a few. And because there are so many sources and more, and that the data is more complicated than before, we will find that the basic business intelligence tools that were used in the past are not able to handle all of the data. This led to the need for more complex and advanced tools to help us to process and analyze the data in the hopes of getting some of the insights and patterns out of the process.

A Look at Data Science

From here, we need to spend some time taking a look at data science even more and see what we are able to do to make this work for our needs. The use of the term data science is going to creep up in a lot of the different things that you are going to try and do with your business. And it is likely that you have heard a lot about this option from your friends and other business partners. But it is important to know exactly what this all means and why it is so important before we dive in.

For example, it is a good idea to have certain skills ahead of time to turn into data scientists. You may want to know what some of the differences are between these data science processes and business intelligence in the past. And then we may want to know how some of the predictions and decisions are going to be made in data science in the first place.

These are just a few of the questions that we are going to have when it is time to work with data science, but first, we need to take a look at some of the different things that we need to take a look at when it comes to data science. To start with, data science is going to be a blend of a lot of different principles of machine learning, algorithms,

and tools that all come together to help us discover what patterns and insights are going to be found in the raw data that we are working with.

For the most part, a data analyst is going to be someone who is going to explain what is going on by processing the history of all the data that we have. On the other hand, we will find that a data scientist is not only going to do the exploratory analysis to help us discover some of the insights that are in it, but it will also use some of the algorithms of machine learning that are more advanced, and some of the basic ones, to help us identify the occurrence of a particular event in the future. But you will find that the data scientist will look at the data from a lot of different angles, and sometimes these angles will not be ones that are known earlier.

So, for the most part, you will find that data science is going to be used in most cases to help us make predictions and decisions. There are a few options that we are able to work with to make this happen and these include options like machine learning, prescriptive analytics, and predictive causal analytics. Let's take a look at how we are able to work with all of these and how we can make sure that they give us the results that we need.

The first option is going to be predictive causal analytics. If you would like to work with a model that is good at making predictions of what is going to happen in the future, you need to apply this kind of analytics along the way. So, if your business is going to provide money on credit, then the probability of a customer making their credit payments in the future on time is a matter of concern for you. You want to make sure that you will actually earn the payments that you deserve, rather than handing over the money and never seeing it again.

With the predictive causal analytics, you will be able to build a model that is able to perform these analytics on how well the customer has paid their bills in the past, and how well they will be able to make the payments in the future. If the numbers look good, then you will be

able to give them the money with a good degree of certainty in order to get started with this process and will loan them money.

Then you are able to work with what is known as prescriptive analytics. If you are looking to work with a model that has the intelligence of taking on its own decisions, and the ability to help modify it with the dynamic parameters, you will need to work with this kind of analytics for it to get this done. This is going to be a relatively new field is all about providing advice. In some other terms, it is going to come in to make a prediction but will suggest a range of the actions and outcomes that are associated with it.

The third option that we are able to work with is going to be machine learning. We can use this in a few different manners, such as making predictions. If you are working on something like the transactional data for a financial company, and you are hoping to build up a model to help determine the future trend, then machine learning algorithms are going to be one of the best bets. This is going to fall under the idea of supervised learning. This is because supervised because you can hold onto the data based on which you can train the machines. For example, you can train up a fraud detection model that can be trained with the help of a historical record of fraudulent purchases to help keep you safe.

There is another method that we are able to work with when it comes to machine learning, and this is going to be to help us with some pattern discovery. If you do not already have some parameters that you would like to use to help base your predictions, then it is important to take some time to find the patterns that are hidden inside of the set of data that you have in order to make the predictions that will move your business forward.

This one, in particular, is going to be more of an unsupervised model because you are not going to have any kind of labels ahead of time for the grouping. There are a number of these algorithms that we are able to work with here, but one of the most common options that will

help us with some pattern discovery is going to be known as clustering.

Data Science and Business Intelligence



If your business has been working with data for some time, it is likely that you have spent some time working with business intelligence at least a little bit. These ideas are going to sound like they are similar, but we are going to take a look at how these are going to have some parts that are a little bit different, and why this is going to be so important for your work.

The first thing that we need to take a look at is going to be known as business intelligence, or BI. It is going to help us to analyze the data that we already have in order to find some hindsight and insight to describe the trends in the business. This one is going to help us to take data from both internal and external sources, prepare it the right way, run queries on it, and then create a nice dashboard to answer the questions that you have. You could work with things like business

problems or work on a quarterly revenue analysis on this one for example. It is also helpful when it is time to evaluate the impact of certain events in the near future.

Then we are able to work with data science. This one is going to be an approach that looks more towards the future compared to BI. This one is going to spend more time exploring things with the focus on analyzing the past or the current data and then using this to help us predict the outcomes that are going to happen in the future. And the main point of working with this is to help us to make some informed decisions in the future. It is a good way for us to learn how to answer the how and the what of the events that are occurring around us.

One of the common mistakes that are often made when it comes to working on projects in data science is that they rush through data collection and analysis, without understanding the requirements, or even having their business problem framed in the proper manner. This is why it is important for us to be able to go through the proper steps, or the lifecycle of data science, to make sure that the project is going to run in the smooth manner that we would like along the way.

The Lifecycle of Data Science

While we are going through this process, we have to remember that there is actually a type of lifecycle that we need to work with. All of these steps are going to be important to some of the work that we are trying to do along the way and missing out on any of them is going to be a bad idea when you want to get the right predictions and insights out of that data. We are going to take a closer look at some of these later on, but you will find that data science has a number of steps that are going to be important when it is time for us to learn the real patterns and insights that are inside of all that data.

The first step in this lifecycle that we need to take a look at is the discovery phase. Before you get started with the project, it is important to understand some of the things that you would like to see happen. It is important to know the specifications, priorities,

requirements, and even the budget that you need to be able to stick with during this project.

You also need to have a good ability to go through and ask the right questions through all of this. Here, you are going to assess if you have the right resources present for your work in terms of having the right kinds of people who can gather and analyze the data, the technology that can get this done, the right amount of time to take on this kind of project, and even the right kind of data to help support the project that we need to get it all done.

During this phase, we also need to spend some time framing what our business problem is all about so that we can get it all done. If we have no idea what our business process is all about, and what problem we want to be able to solve, then with all of the data that is there, we are going to end up with a big mess and will spend way time and energy when it is time to look through that data. And this can help us to formulate some of the hypotheses that we would like to get started and to help us test things out as we go through it.

Then we move on to the data preparation part of this process. In this kind of phase, we are going to need to work with what is known as an analytical sandbox to help us get the analytics that we would like done for the entirety of the project that we are doing. This is going to make it easier for us to go through and explore some of the different parts that are present in our work and will ensure that we are going to really get a chance to learn about the data that is there.

During this phase, we need to spend some time learning more about the data. This means that we need to work on exploring, preprocessing, ETLT, and to help us get data into the sandbox that we are working with. This is all-important to ensure that we are set up with and will ensure that we are going to really get set up here.

But before we are able to put the data through the model that we are hoping to use, later on, you will find that the best step that we have to take is to go through the data and really prepare it. If there are

missing values, duplicate values, or it is not in the right format, then you are going to end up in trouble and the algorithm is either not going to offer you any kind of results, or the results that you get will not be as accurate as you would like.

There are a few languages that you are able to use when it is time to clean, transform, and visualize the data that you are working with. You will find that this will work well if you focus on the R or the Python language. This is going to be a good step where we get to spend time on outliers in the data and will help us to establish the relationship that we need between the variables that we have. once you have been able to clean up and then prepare the data, it is then time to move through and do what is known as exploratory analytics on it to get the best results.

The third step that we are able to work with is going to be known as model planning. In this one, we are going to spend some time determining which techniques and methods we would like to use in order to draw up the relationships that we need with variables. These relationships are going to set the base for all of the algorithms that you would like to implement in the phases that we will get to later.

For this step, we need to spend some time applying the EDA, or working with Exploratory Data Analytics with the help of a lot of different statistical formulas and visualization tools. There are a lot of options that you are able to work with when it is time to bring out the model planning tools, and some of the options that you are able to work with will include things like the right coding language, the SQL analysis services, and even the use of SAS and ACCESS.

Although there are many tools that are on the market for you to work with, you will find that the Python language is going to really work well for helping you to get through all of the processes that we are going through. You need to make sure though that you are able to spend some time planning out the model that you would like to use in order to get all of this done at the right time. There is the possibility that you will come up with more than one option when it is time to

prepare a model, but we need to learn more about these and figuring out how we are able to work with each one, so that we can pick up the model and the algorithm in order to get the results that we want in the end.

From this point, we need to spend some time looking at the process of model building. When we are on this kind of phase, we are going to spend some time developing the sets of data that we would like to work with training and testing purposes. We can't just grab a model and assume that it is going to work that we would like to see. The results will not be accurate because the model as of yet, has no idea what you are hoping to see in the process.

Instead, you need to spend some time looking at the process that is necessary in order to really train the data that you would like through the model. This will ensure that it knows what you should work with, and over time, we will find that this is going to work better than ever before. But the right sets of data for training and testing will be necessary before you can rely on the patterns and insights that you will need.

During this phase, we really need to consider whether the tools that we already have will be enough to run the models that we have, or if we would need to bring in some more of a robust environment. You will analyze a lot of the different learning techniques that are needed in machine learning in order to help clustering, association, and classification to help up the model building that we would like to do.

The fifth step that we are going to focus on is going to be known as the operationalization. When we get to this phase, you are able to deliver some of the different parts that we need to explain what you were able to find in all of the models and algorithms that you have run that data through in the previous step. This one is going to include providing the people in the business who make big decisions information through technical documents, reports, code, and briefings.

In addition, you will find that in this step, it is possible to take some of the information that we have collected and used it in a type of pilot project that is done in real-time. This will usually only take place in a small part of the business, such as one of the offices, in order to see how it will work and whether it is the right option for you. If it works, then it can be expanded out to other parts of the company. But if it ends up not working, then there isn't as much waste in time or money to make this happen.

The pilot project is going to be important because it is going to really help us to see how the information that comes from the data that you collected is going to work, and whether it is the right decision for your business, without as much risk. It is a great way to get a clear picture of the performance and some of the other constraints that are related here on a small scale before we go through and work with a full deployment.

And finally, we are going to end with some communication of the results that we have. When we get to this point, we are going to find that it is important to evaluate whether or not we have really been able to achieve the goals that we planned out in the first phase. So, with this kind of phase, we are going to start by taking all of the key findings that we had in all of the other steps, and then organize it in a manner that will make it easier to understand.

This is often going to be done with the help of reports, spreadsheets, and other visuals that will help others who may not have the added technical experience in order to understand what is going on in the long run either. The data scientist has to go through and really go through this information and make sure that they are presenting it in a manner that the decision maker's in the company will understand. The methods that they will use depend on who they are working with.

So, once you have all of these key findings in place, you will need to communicate the information to the right places and then determine if the results of any pilot projects that you do will be a failure or a

success based on the different parts and criteria that you were able to develop at the beginning of this project.

There are so many times when we are able to work with the data analysis and we will be able to see what data science is able to do for our business. But we have to make sure that we are going through and writing it out in the right manner, and that we use the right methods and models in order to get it done. We will explore through more of this as we go through this guidebook, but you will find that these steps above will help us to get more done in the process as well.

Chapter 2: How Does the Python Language Fit In With Data Science?



Now that we have spent some time taking a look at what data science is all about, it is time to bring in the work that we are able to

do when it comes to the Python language. When we work with these models and all that can come up with data science, you will find that we need to write out some algorithms. We will take a look at machine learning later on, but Python will work well with machine learning and can help us to write up and execute the algorithms that we want to create in this process as well. We could go with some of the other languages that are out there, but Python is often one of the best options to use.

The Python language is going to be one of the best options to work with when it is time to work with any of the projects in data science that we would like to accomplish. There are actually going to be a lot of projects that you are able to do with these ideas of data science, and Python is able to help us to get more of it done in the process. There are also a lot of other reasons that you would want to work with the Python language in order to make sure that you get your projects done.

This language, even though you are going to find that there are a number of complexities that are able to come with it as well, is going to be really easy for a beginner to learn more about. If you are someone who is just getting started with coding to help you finish the algorithms that we have in this, and you are a bit worried about getting started, you may find that Python is going to be the best one to work choose from. Python is designed for the beginner to work with, and even if you have never done anything with the coding of any sort in the past, Python will make the process as simple and painless as possible.

One of the main points of choosing Python for some of your coding needs is that it makes sure that your coding is as simple and easy as possible, whether you are a professional coder or someone just starting out. The words that are used in this are English, the syntax is simple, and it relies on the idea of being an object-oriented programming language. This means that it is easy and powerful and will meet your needs when it comes to finishing up a data science kind of project.

There is still a lot of power that comes with the Python language, even though it is designed for a beginner to learn to code for the first time. Many people worry that when they get started with the Python language that it is going to be too simplistic. They may reason that because this kind of language is designed to help beginners get started with their work, it is going to be too simple in order to get any of the more complex codes done and ready to go.

This can't be further from the truth. You will find that working with Python is a great option, even with the ease of use, that is strong enough to handle some of the more complex codes and projects that you would like to get done. For example, Python is able to help out, with the right libraries and extensions, things like machine learning, deep learning, science, mathematics, and other complicated processes, whether they are needed for data science or not.

More productivity for the programmer: The Python language has a lot of designs that are object-oriented and a lot of support libraries. Because of all these resources and how easy it is to use the program, the programmer is going to increase their productivity. This can even be used to help improve the productivity of the programmer while using languages like C#, C++, C, Perl, VB, and even Java.

When it is time for you to work on some of the different parts of your data science project, having more of that productivity is going to be so important overall. You will find that when the programmer is able to get more things done in a shorter amount of time, it ensures that they are going to see a big difference in the project they are working with.

Integration features: Python can be great because it integrates what is known as the Enterprise Application Integration. This really helps with a lot of the different things you want to work on in Python including COBRA, COM, and more. It also has some powerful control capabilities as it calls directly through Java, C++, and C. Python also has the ability to process XML and other markup languages because

it can run all of the modern operating systems, including Windows, Mac OS X, and Linux through the same kind of byte code.

While we have spent most of our time right now focusing mainly on how we are able to work with Python in order to make sure we finish up any of the projects that we want in data science, there are going to be a few data science libraries that we have to add in with Python to gain the kind of compatibility that we need and to make sure that we are able to properly handle our algorithms and get the work done.



But this is part of why working with Python here is going to be so amazing. Python is compatible with many language options, which means that you can use a lot of the libraries and extensions that you want for a project, while still getting some of the ease of use that you will need from Python.

Another benefit that we are able to focus on here is that Python has a very large community. For someone who is just getting started with coding, having a nice big community to answer your questions, to help show you the best way to get started, and more will really be

helpful. Data science and some of the algorithms that are needed for it are going to really need some complex coding, and Python is going to have a nice community to help you out with this.

The community that is available with Python comes from all parts of the world, and you will find programmers of all different coding levels. They can offer you some advice, give you some of the codes that you need, and will make it easier to get through some of the issues that you may face when it comes to some of the algorithms that you have in data science.

The next benefit on the list that we get to enjoy is some of the standard libraries of Python. This library is going to come with a lot of power to make sure that your coding tasks get all done. When you download this language at the start, you will find that this library is going to come with it, and will already be able to handle a lot of the functions, methods, and codes that you would like to do right from the beginning.

Keep in mind though that while the standard library of Python is going to be powerful and provide you with lots of options, when it comes to working with data science, there are going to be times when you need to add on some extensions to help make sure that you can get machine learning and more done. But there are a lot of choices when it comes to these, and learning how to make them work will make a difference in some of the coding that you would like.

Simply by working with the standard Python library that comes with your installation, there are a lot of powerful types of codes that you are able to write out including conditional statements, inheritances, and loops. All of these are things that you are able to use, and even though they are basics of learning to code, they will help out with some of the algorithms that we want to use later on.

There are a lot of special extensions and libraries that we are able to do with Python that is perfect for data science and machine learning. Even though there is a lot that we are able to do when it comes to

using the Python language to help out just with the standard library, there are also a number of libraries that work not only with Python, but also help us to get more done with data science at the same time.

We are going to take some time to look at a lot of the different libraries for data science and Python, and you will get a chance to see how well these can help out with some of your data science projects. Some of the best options though, based on which machine learning algorithms you want to work with, include TensorFlow, Scikit-Learn, NumPy, SciPy, and Pandas.

Python is one of the options that is used to help out with creating the machine learning algorithms that you need to create your own models with data science. Without these models, you are going to end up having trouble going through all of that data you have been able to collect and actually find the insights and predictions that you want from that data. There are other languages you can use, most notably the R programming language, but Python is one of the best to help get all of this done.

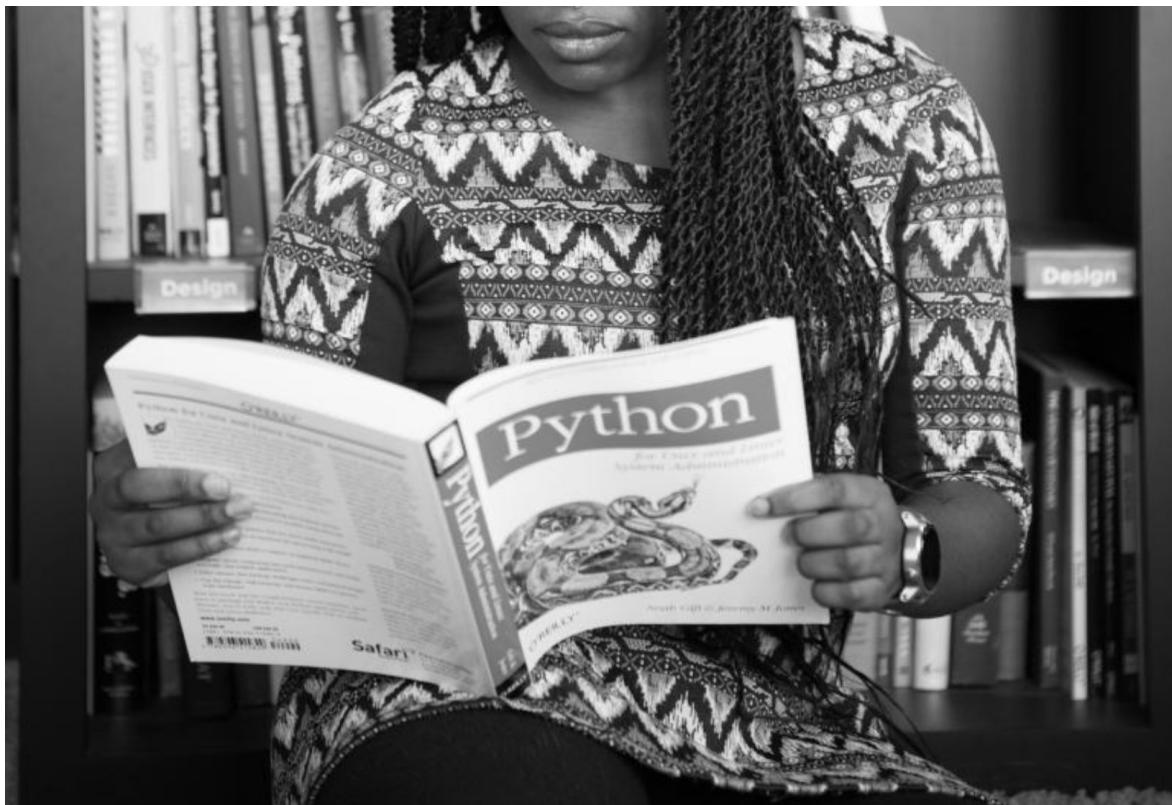
There are also going to be a few steps that we are able to focus on when it is time to work with the projects that we need in data science. But one of these parts, the analysis part, we will have to work with a number of techniques including machine learning and deep learning. These help us to create a model that is needed in order to handle the large data that we want to work with. Going through all of this data is something that is pretty much impossible when you work on it in a manual manner. but with the right model that is run by the Python language, you will find that it is easy enough to sort through all of that information and get the predictions and insights that you are looking for out of the data in no time at all.

And this is really one of the main reasons that people will choose to work with the process of data science in the first place. The companies that decide to go with this because it allows them to take a large amount of raw data, and then figure out the insights that are

found inside of that data. The models that are created in order to get this done, thanks the Python language and machine learning, can really make it easier for us to meet these goals and will make it easier to get ahead of the competition, meet the needs of the customers and the industry, and so much more.

There are going to be a ton of reasons why a data scientist is going to work with the process of this analysis with the help of the Python language. There are options in other coding languages that we are able to use, but you will quickly see that the Python language is going to be one of the best options that we can focus on to get the models created that we want, to work with the right kinds of libraries that we want with the use of machine learning and the right algorithms, and so much more.

Chapter 3: The Best Python Libraries to Help with Data Science



While we are on this topic of how well Python is able to work with data science, it is time to take a look at some of the different libraries that you are able to add onto Python and use when you would like to make sure that this language is going to work with some of the projects that you would like to accomplish with the data science project that you have in mind.

Remember that we talked a bit before how the standard library that comes with Python and how it is going to help us to learn a lot of the coding that we need and get things done in no time. however, we will find that a few of the models and algorithms that we want to do with data science are going to need a little bit more when it is time to work with your projects rather than the standard library.

The good news here is that we are able to work with some of the different libraries that are out there in order to help us get this done. And Python is going to have quite a few libraries and extensions that we are able to work with that can handle or data science projects and the other models and algorithms that we want to be able to work with overall. We just need to make sure that we go through and pick out the one that is going to work for the specific model or algorithm that you have in mind in the first place.

All of the libraries that you want to work with here are going to be a bit different and will handle some of the work that you want to do in a different manner. Some of them are best for helping with the analysis and some may be better for handling some of the data gatherings that you want to do. This is why we need to learn a bit more about some of the libraries and what they will do for our needs.

These libraries are so important when it comes to some of the work that you want to do with algorithms and models. Without these libraries, you will not be able to complete some of the necessary tasks that come with data science, and it is going to really make it almost impossible to work with the predictions and insights that we need out of all there.

There are quite a few libraries that not only work with the Python language but will work with machine learning, data science, deep learning, and so much more. Some of the different libraries that you are able to pick from to get all of this done will include the following:

NumPy

When we first get started with doing some data science on Python, one of the best libraries to download is going to be NumPy. Many of the other data science libraries are going to rely on some of the capabilities that come with this kind of library, so having it set up and ready to go on your computer is going to make a big difference.

When you are ready to start working with some of the scientific tasks with Python, you are going to need to work with the Python SciPy

Stack. This is going to be a collection of software that is specifically designed to help us complete some of the scientific computing that we need to do with Python. Keep in mind that this SciPy stack is not going to be the same thing as the SciPy library though so keep the two of these apart. The stack is going to be pretty big because there are more than 12 libraries that are found inside of it, and we want to put a focal point on the core package, particularly the most essential ones that help with data science.

The most fundamental package around which this computation stack is going to be built around is NumPy, which is going to stand for Numerical Python. It is going to provide us with an abundance of useful features for operations that you want to handle with matrices and n-arrays. This library is going to help us with a lot of different tasks that we want to do, including the vectorization of mathematical operations on the NumPy array type, which is going to ameliorate the performance and will speed up the execution that we see at the same time.

SciPy

Another library that we are able to take a look at when it comes to working with the Python language is going to be SciPy. This is going to be a library of software that we can use to help us handle some of the tasks that we need for engineering and science. If this is something that your project is going to need to spend some time on, then SciPy is the best library to get it done. You will quickly find that this library is going to contain some of the different modules that we need in order to help out with optimization, integration, statistics, and even some linear algebra if we would like to name a few of the different tasks that work well with this.

The main thing that we will use with this library and some of the functionality that you will need when bringing it up is that it is something we can build up with the help of the NumPy library from before. This means that the arrays that we want to use in SciPy are provided to us thanks to the NumPy library.

This library is going to provide us with some of the most efficient numerical routines as well as some of the numerical integrations that we need, the help of optimization, and a lot of the other options that we need with our specific submodules. The functions that are going to be discussed in this library are going to be documented as well to make it easier.

Pandas

We can't go far in our discussion over the libraries in Python that work with data analysis without spending some time looking at the Pandas library. This one is going to be designed to help us out with all of the different steps that we need with data science, such as collecting the data, sorting it and cleaning it off, and processing the various data points that we are working with as well. We are even able to take it a bit further and look at some of the visualizations that are needed to help showcase the data in a manner that is easier to work with.

The Pandas library is going to be a package that will come with Python and has been designed so that it can specifically make some of the work that we need with labeled and relational data simple and more intuitive. Pandas are going to be the best tool that we can use to help out with many of the processes that we want to handle, and this can include some of the data wranglings that needs to happen in this process.

In addition to some of the benefits that we have talked about before, the Pandas library is going to work well when it comes to easy and quick data visualization, manipulation, and aggregation, along with some of the other tasks that we need to work within order to help us get our work done in data science.

Matplotlib

As we are working through some of the libraries and projects that we want to focus on with data science, we are going to find that working with some data visuals can be helpful as well. These visuals are going to make it easier for us to handle the complex relationships that are found in our information and our data in the first place.

For most people, it is a lot easier to go through and understand the information that we have when it comes to some sort of visual, whether this is in a picture, in a graph or chart, or some other method. At least compared to some of the methods that we can use with reports and spreadsheets. This is why the visualization process of data is so important when it is time to work with data science. And this is why we need to look at Matplotlib to help us to take care of these visuals.

Matplotlib is going to be one of the best data science and Python libraries to work with to make sure that we can create and handle some of the simple and most powerful visuals in no time. It is going to be a really strong piece of software that will help us to take the results that we are getting when we do the algorithms, and then effortlessly turning them into something that we are able to see and understand easier than before.

We have to remember here that when we are working with the matplotlib, you will find that it is going to be low-level. This means that you are going to need to spend more time writing out more code to help all of this get done and to give us some of the higher-levels of visuals that we would like. It requires a bit more effort than we are maybe used to working with, but it is going to still provide us with some of the things that we need to get our work done. Just be aware that it does require some more work.

When we are working with this kind of library, we have to look at it to help us see how we are able to handle pretty much any kind of visual that we would like. But we have to remember that we are working with a lot of data and go through the algorithms to understand that information first. Some of the different options that you are able to work with when it comes to these visuals though will include the following:

- The step plot
- Contour plots

- Quiver plots
- Spectrograms
- Pie charts
- Histograms
- Bar charts
- Scatter plots
- Line plots

In addition to helping you to work through some of the different plots and graphs that we have above, it is possible to work with a few of the other capabilities that happen with this language and this library. You can use this kind of library, and some of the features that we need, in order to work with creating grids, legends, and labels to make the formatting of our visuals easier to handle. There is a lot that we are going to enjoy when it is time to handle these visuals with matplotlib, and it is definitely an option that you will want to spend some of your time on.

Scikit-Learn

This is going to be an additional package that you are able to get along with the SciPy Stack that we talked about earlier on. This one was designed to help us out with a few specific functions, like image processing and facilitation of machine learning. When it comes to the latter of the two, one of the most prominent is going to be this library compared to all of the others. It is also one that is built on SciPy and will make a lot of use on a regular basis of the math operations that come with SciPy as well.

This package is a good one to work with because it can expose a concise and consistent interface that programmers are able to use when it is time to work with the ones that go with the most common algorithms of machine learning. This is going to make it simple to bring machine learning into the production system. The library is able to combine together quality code and good documentation, which can bring together high performance and ease of use, and it is one of the industry standards when it comes to doing anything that you need with machine learning in Python.

Theano

We can also spend some time working with the Theano library, and we will find how this one is going to work the best when we want to handle more of the deep learning process rather than machine learning like the other options. This library is going to be a kind of package from Python that is able to handle arrays that are more multi-dimensional, similar to what we saw with NumPy library and some of the mathematical expressions and operations.

When we work with the Theano library and we get it all compiled, which means that we get it to run as efficiently as possible on all of the architectures along the way, it is going to help us to get so much done in no time at all. This library is going to be so great with some of the deep learning that we want to accomplish, and it is worth our time if we want to focus more on the deep learning that we need.

One of the most important things that we will be able to focus on when it comes to working with the Theano library is that it is really great at integrating tightly with the NumPy library on some of the operations that are considered lower in level. The library is going to help us to optimize any of the GPU and CPU that you are working with, which is going to help us to go through these computations faster than before. Add in that this library is going to be more efficient and stable and you will get precision in your results that weren't possible in the past, and you will see why this is a great option to go with.

TensorFlow

The next library on the list that we are able to talk about is going to be known as the TensorFlow library. This is going to be a library that is special because it was originally developed by Google and it is also going to be open-sourced so that we are able to use it for our own needs in no time. It also comes with computations for data flow graphs and more that have been sharpened in order to make sure that we can handle machine learning.

In addition, we are going to find that this library is going to be one of the best to choose when it is time to work with neural networks. These networks are a great type of algorithm to handle because they will help us to handle our data and make some good decisions through the system. However, we have to remember that this is not something that is only specific to Google's company. It is going to have enough power behind it and will be general-purpose enough to help us out with some applications that are better for the real-world.

One of the biggest features that we are going to need to focus on when it comes to this kind of library is that we are likely to see a lot of nodes that are in many layers when we work with the system. This is going to be great to work with because it will help us to train any of the artificial neural networks that we have, even when we have a set of data that is really large. This is going to make it easier to handle some of the models and algorithms that we are looking to create. For example, this is a library that has been used to help with voice recognition and even the identification of objects in a picture that is presented. And these are just a few of the options that we will be able to see with this kind of library.

Keras

And the final library that we are going to take a look at in this guidebook is the Keras library. This is going to be a great open-sourced library that is going to help again with some of the neural networks that we want to handle in this language, especially the ones that happen at a higher level, and it is also written in Python to make things easier. We will find that when it comes to the Keras library, the whole thing is pretty easy to work with and minimalistic, with some high-level extensibility to help us out. It is going to use the TensorFlow or Theano libraries as the back end, but right now Microsoft is working to integrate it with CNTK as a new back end to give us some more options.

Many users are going to enjoy some of the minimalistic design that comes with Keras. In fact, this kind of design is aimed at making our experimentation as easy and fast as well, because the systems that

you will use will still stay compact. In addition, we will find that Keras is going to be an easy language to get started with, and it can make some of the prototyping that we want to handle easier.

We will also find that the Keras library is going to be written out in pure Python, and it is going to be a higher level just by nature, helping us to get more programming and machine learning done on our own. It is also highly extendable and modular. Despite the ease of using this library, the simplicity that comes with it, and the high-level orientation, Keras is still going to have enough power to help us get a lot of serious modeling.

The general idea that is going to come with Keras is based on lots of layers, and then everything else that you need for that model is going to be built around all of the layers. The data is going to be prepared in tensors. The first layer that comes with this is then responsible for the input of those tensors. Then the last layer however many layers this may be down the road, is going to be responsible for the output. We will find that all of the other parts of the model are going to be built in between on this to help us get the results that we would like.

As we are able to see through this chapter is that there are a ton of libraries that we are able to work in order to help us out with Python and will help us get things done when handling machine learning and all of the other parts that we want with data science. Creating some of the models that come with machine learning, and making sure that the data we have collected in the raw form is actually able to be changed around to make sense and help us to make good business decisions is so much more successful when we are able to work with some of these Python libraries to get it done.

All of the different libraries that we have spent some of our time discussing and learning about in this guidebook will be able to help us handle a lot of the different parts of the process, and will help us out in a lot of manners as well along the way. It is important as someone or some company who wants to work with data science to make sure that we are going with the best library for what kind of

project that we want to handle. And you will find that all of them can help us get things done and will provide us with a lot of the tools that we are looking for as well.

Chapter 4: Gathering Your Data



The first thing that we need to take a look at is how to gather up the data that we need to accomplish this kind of process in data science. We need to have a chance to go through and look at our data, figure out what kind of data is out there that we can use, and so much more. But figuring out where to get that data, how much to collect, and what kind is going to be right to help us figure out more about our customers and industry, can be hard.

There is an overabundance of options out there when it comes to the kind of data that we want to use along the way. We need to make sure that we are picking out the right kinds of data, rather than just

collecting data because it is there and looks like the right thing to work with. When we are able to organize this in the manner that we need, and we make sure that we actually get the good data, even if it is not organized and structured the way that we want in the beginning, it is going to be so important.

That is why we are going to spend some time in this chapter exploring what we are able to do with our data, how it is going to work for our needs and even some of the places where you are able to look in order to find the data that you would like to work with. With that in mind, we need to dive right in!

Know Your Biggest Business Problem



There is a lot of data out there, and it is not going to take long doing some searching before you find that you will end up in a rabbit hole with all of this information if you don't have a plan or a direction for what you are going to do with all of that information. There is a ton of

good data, but if you just let it lead you rather than having a clear path in front of you, you are going to end up with a lot of problems and will never get the decision making help that you need.

If you have already gathered up your data, then this point is gone and we just need to work from there. You can form through your biggest business problem, the one that you would like to spend your time focusing on and fixing, and then sort through the data there and see what changes you are able to make and what data out of that large source you have is going to make the biggest difference. Don't be scared to just leave some of the data for later, and don't let the fact that you may not use some of the data hold you back either.

During this time, we want to focus on knowing the best information, this is going to be the best way to make sure that you get the information needed to really propel your business into the future. Even some of the information is left behind, that is fine. You may come back to it later if you need some of it. But only the best data that you have should be used for your algorithms to give you the best results.

Now, if you have not had the time to go and collect any data yet, this is something we can work with as well. Forming the problem that you would like to solve, and having a clear path can help you to sort through all of the noise that is out there, and will ensure that you are really able to get things done in the process. You need to make sure that you are searching in the right places, and looking for the information that is going to be the most critical for what you are trying to accomplish, the part that is going to be so important when it is time to handle some of the work that is out there.

Places to Look for the Data

The next thing that we need to consider when it comes to this process of gathering up the data and using it in the manner that we would like is figuring out where to find and look for the data that we need. There are actually so many different places where we are able

to look for the data that we want to handle, but this is part of the beauty of the modern system that we are using today.

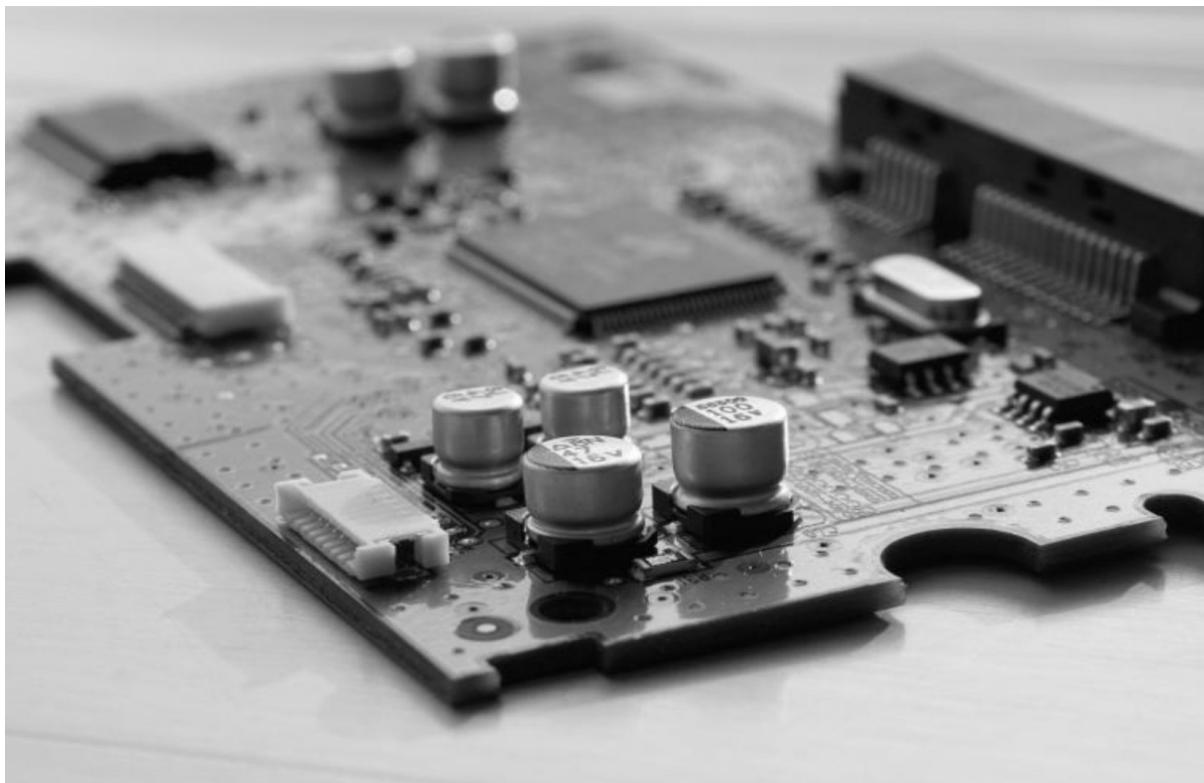
We have to remember though that most of the data that we will collect today is not going to be organized or structured. We will look at some of the steps that you are able to take in order to organize the data a little bit later, so this is not a big deal. Just be prepared that you will have to go through and take on a few extra steps in order to make sure that your data stays organized in the manner that you would like and that it is not going to be as nice and neat as you would like in the long run.

So, the places where you are able to look for some of the data that you would like to use in this process will be varied and it often depends on what you are hoping to get out of this process. You want to concentrate on getting the highest-quality data in the process that you can though. This is going to ensure that you are going to be able to find the data that you need and that the algorithms you use later on will really be able to provide you with some of the best results and insights that you need to move your business forward.

There are still a lot of places where you are able to look to find the data that you want. You will find that you can pick out data from websites (especially if you would like to work with web scraping), from social media sites if you are using one from surveys and focus groups of your own, and from other companies who may have collected the information and are using it to help out others along the way.

You may find that if you are able to bring up data from a more unique source as well, this is going to get you even further ahead with some of the work that you want to do. It will ensure that you will have data that no one else is going to have, and will provide you with some new patterns and insights, as long as you make sure that the data is high quality and will actually be good for your needs.

Where to Store the Data?



We also need to consider where we would like to store some of the data that we are working with along the way. You are likely to gather up a lot of data in the process, and it isn't likely that you just want to have it sitting around without a purpose or having it in a safe and secure location. This is especially true if you are working with data that is your own, data you got from surveys and other places that you don't want others getting ahold of.

There are a number of different places where you are able to store this data for your own needs and the location that you choose is often going to depend on what works for you. If you have enough storage space on your own network, this can be a great place to start. Then the data is always safe and secure with you and easy to reach. You just need to make sure that you are keeping some good security measures on your system so you don't end up losing that information and no longer having it at your disposal.

Many companies decide to put it on a web-based storage area, like the cloud. This adds in another level of protection to the information and will ensure that you are able to reach that data when you need it as well. There are a lot of these kinds of storage areas that we are able to work with, and you will find that you are able to get this to work for some of your needs pretty well. Whether your storage needs are large or not, you will find that storing this data is going to make a world of difference when it is time to handle this process, and you just have to decide how much you would like to use ahead of time.

Knowing where to find the data that you need to start out with your data analysis and data science project is going to be super important. This is going to set the tone for the work that you are able to do later on and how much success you are going to have with your project as well. Make sure to search around for the data that is going to be needed in this, and pay attention to how much of it you will need, where you are likely to find it, and more.

Chapter 5: Organizing and Cleaning the Data



As you get into the process of working with data science, you will quickly find that the vast majority of your time with a project is going to be spent collecting, cleaning, and then preparing the data that you have to work in analysis. This is because many times, the sets of data that we have are going in a lot of sizes, formats, and more, and if they are not organized and ready to go, you are going to run into some problems with getting the project to work, and ensuring that you get accurate results.

This is why we will want to spend some time cleaning and organizing the data that we are working with along the way. There are a number of methods that you are able to use to make this work for your needs. But in the end we want to make sure that we have the data organized, usually in a database of some kind, the duplicates handled and gotten rid of, and we need to have a plan for some of the missing values and outliers that are found in your data. Let's dive into some of the basics of data preparation and why this is so important to some of the work that you need to do in a data science project.

What is Data Preparation?

Let's say that we are going through and trying to get a good analysis of the log files that are on a website so that we can figure out which IP address a spammer is coming from. You can also use this to figure out which demographics your website is reaching and getting more sales with, or which region geographically your website is the most popular in. What steps would we need to take in order to figure these things out?

To help us to answer all of these kinds of questions, we need to perform an analysis of the data and we need to do this with a few important columns. This is going to include the number of hits that a website gets, and the IP address of the hit. As we know, the log files that we may use are not going to be structured and this means that we will have to go through and sort out the unstructured information as best we can. This requires some work and some good organization and cleaning in order to make it all work out.

The idea of data preparation is where we are able to take all of that data that is kind of a mess and doesn't have the formatting and more that we want, and we turn it into a form that is easy to use and will flow through our chosen algorithms later on if we would like. This does require a number of steps in order to be successful, and often it is not fun.

For example, there are a lot of studies out there to look at this part of the data science project, and it is going to be really hard to miss out on this step. In fact, it is estimated that when it comes to a data science project, you will spend up to 60 percent of your time organizing and cleaning the data. This is compared to all of the other steps having to take up other parts of the process, and it is pretty amazing how important this process is.

The reason that this process is so important is that it needs to have clean and organized data in order to get things done. If the data is a mess, and if there are a lot of outliers or information that does not match up the way that it should, it is going to be a waste of time. The

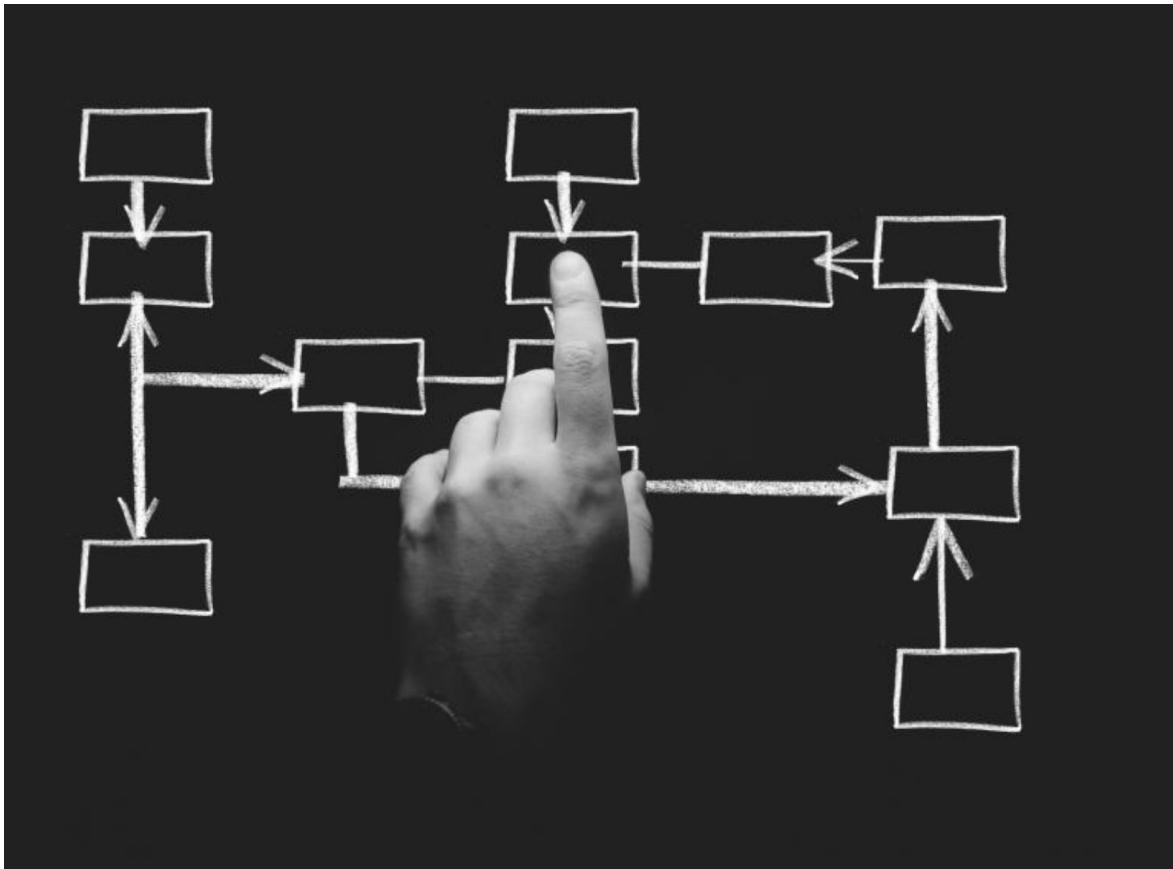
algorithm will not be able to go through the process in the right manner, and it is going to provide you with answers that are not accurate.

For example, if you go through and find that there are a lot of missing or duplicate values, then this can really skew the information that you are getting out of this information. If there are a lot of duplicates, then the information and the results will start to skew a bit towards this one as well. On the other hand, if you end up with a good average, but then you see that there are a few outliers that really are far away from this average, it could mess with some of the results as well.

This is why cleaning up the data is going to be so important. We want to make sure that the results we are putting into the algorithms, and the outputs that it gives out to us in the end, are going to be accurate. Many companies are going to base a lot of business decisions on these results, so the more time that we can spend cleaning that data and keeping it as organized and easy to use as possible is going to ensure that the algorithms work the way that we want.

However, you will find that even though the cleaning and preparation process is going to be so important, it is not much fun. About 57 percent of data scientists out there find that cleaning and organizing the data, even when it is so important, is going to be one of the most boring and least enjoyable tasks that come with this process.

Why Is Data Preparation So Important?



To see why this is so important, we need to take a look at an example of how this can work. Let's say that you are working as a data scientist and you are going to need to estimate how many burgers your store is going to sell on a daily basis. You are going to work with a file that will have rows that will describe some of the financial information about the system. There are going to also be some columns like state, city, and how many of the burgers are sold.

However, rather than having all of this data show up in one document, it is likely that information is going to be split up in a lot of different files, and most of them are going to come in a lot of different formats so it is difficult to look it all over or combine them with one another.

This is the job of a data scientist. They will need to join together all of the data and make sure that the combinations they get make sense so they can continue on with the analysis. Usually, there are going to

be several formatting inconsistencies and floating issues that show up in the set of data. For example, there are going to be some rows where you see that the state is going to be 101 and then the number of burgers could be something like New York. This is a mess, but it is sometimes what happens when we move things around and try to combine them with one another. And that is exactly why we need to work with the process of data cleaning.

Data cleaning is a process that is going to require that the data scientist goes through and finds all of the glitches, fixes them up, and then make sure that the next time this information comes in, it is going to automatically be fixed along the way. predictive analysis results of a data scientist can be as good as the data they have assembled. Data preparation is going to be vital and important to the process that you are doing and can help us to actually find some of the insights that we are looking for out of that data.

With all of the great data that we are able to gather over time, you will find that the problem isn't finding the data that you need. The problem is going to be making sure that the data is organized and ready to work with when you would like. This takes some time, but high-quality data is going to be the backbone of some of the things that you want to do with this process. Making sure that the data is high quality and will work the way that you want is going to be important to this as well.

Steps Involved for Data Preparation

There are a few steps that we are able to work with when it is time to handle this process of data preparation. The first one is going to be where we work on data cleaning. This is where we are going to spend our time correcting some of the data that is not consistent and then filling out some of the missing values while smoothing out some of the data that is making a lot of noise.

Keep in mind with this one that there could be a lot of rows in the set of data that are not going to have value for attributes of interest, or

there could be some data that is not consistent or has duplicate records. And sometimes it is just another random error to work with as well. All of these data quality issues are going to be tackled when we are in this kind of process.

The missing values are going to be tackled in a lot of different manners depending on the requirements that we find. This can happen when we either ignore the tuple or when we work to fill in the value that is missing with the mean value of the attribute. There are a few other options that we are able to use for this one such as the global constant, the Bayesian formula, or the decision tree. Noisy data is something that we are able to tackle manually through a lot of regression or clustering techniques.

The second step that we need to focus on here is going to be known as data integration. This one is going to involve a number of steps to help us get some of the work done that we need including schema integration, resolving data conflicts if there are any there, and then handling redundancies in the data.

The third step that comes up when we are preparing our data is data transformation. This is a step that requires us taking away any of the noise that is found in our data. When that noise is done, we can then work on normalization, aggregation, and then generalization of the data to help us get the results that you would like.

The next thing that we are able to work with is going to be known as data reduction. The data warehouse is going to contain some petabytes of the data and then running the analysis on the complete data that is present in the warehouse. This is going to be a process that is really time-consuming. In this step, the data scientist is going to obtain a reduced representation of the set of data, that is going to be smaller in size, but yields are going to be almost the same as the outcomes of the analysis.

When we are working on this step, you will find that there are a number of data reduction methods that you are able to apply to your

data. The kind that you are going to use will often depend on the requirements that you have with some of the results that you are able to work with this. Some of the data reduction methods that you are able to work with will include numerosity reduction, data cube aggregation, and dimensionality reduction.

And finally, the final step that we are able to work with when it is time to do some of our data cleanings is going to be the data discretization. The set of data that you are going to work with will often contain three-man types of attributes. These attributes are going to include ordinal, nominal, and continuous to talk about a few.

It is possible that some algorithms are going to only go through and accept attributes that are categorical. Data discretization is going to be a step that is good for helping the data scientist to divide up some of their continuous attributes so that they fall in smaller intervals and will make it smaller in size as well. This is one of the best ways to make sure that the data is going to be set up and ready for you to use.

As you can see here, there are going to be a lot of techniques and methods that you are able to use and many more that are developed to help out with the preparation of your data in this stage. But it is still very much in the early stages of what we can do, and many data scientists are still working with it in order to find some new strategies and techniques that they are able to use for some of their own needs.

It is so important for you to spend time learning how to clean off the data that you are working with overall. This is going to make a world of difference in how successful you are going to be with some of your work, and will ensure that the data is going to actually be ready for some of the algorithms that you are going to work with later on.

The data cleaning process is going to take up quite a bit of the time that you need for your data science and data analysis process, even though it may not be as fun to work with as some of the other parts along the way. making sure that you are set up and ready to handle

some of the work that comes with it, and understanding why this process is so important in the first place, is going to be so important to ensure that you get the results that you would like along the way. follow some of the steps that are in this chapter, and you will be able to clean and organize the data in the way that you need, ensuring that all of that data is ready to go when it is time to handle your data analysis.

Chapter 6: A Look at Data Mining



Another topic that we need to spend some time on here is a bit of the work that comes with data mining. This is going to be a new topic that we haven't had a lot of time to talk about yet. But it is going to be a specific part of the data science process that we need to focus on, and we need to take a closer look in order to help us get the data working and performing the way that we want. That is why we are

going to take a look at why this data mining is so important and what we are able to do with it.

The first goal that we want to focus on when we are working with this idea of data mining is to learn more about what this process is. data mining is going to be the steps that a company is going to use in order to take all of the raw data that they have, and turn it around into information that is useful. It is going to be a bit more specialized than the original steps we talked about in the first chapter. But with some of the work that we can do with machine learning, Python, and a good type of software, a company is going to work with data mining to help look for some patterns that will show up in the larger batches of data.

The reason that we want to take some time to look through all of this data and see what is inside of it is that there is often quite a bit of data that we need to go through to start with. Businesses are able to look through a lot of data, and hopefully, if the data mining works well and does what they would like, they will be able to learn something new. Whether they learn how to better serve their customers, how to beat out the competition, and even how to develop some marketing strategies that are better, it is going to help us to increase sales, cut down on costs, and really help to reduce the risks as much as possible.

There are a lot of parts that we have to make sure comes together if we would like to make sure that the process of data mining is going to work the way that we would like. There are a number of steps that we are able to take a look at when it is time to work on the process of data mining, and this is going to include a few things including collecting enough data, warehousing, and processing power on the system of your computer.

The Process of Data Mining

Using some of the information that we have above, and keeping that in mind, we need to spend some time looking at the basics of what

we have with data mining overall. Data mining is going to involve a company going through the process of exploration and then analyzing a large amount of the information, in the hopes of figuring out some of the patterns and trends in the data that are the most important and meaningful. As we can imagine, it is likely that they are going to be able to find a ton of insights and some other information that is valuable inside of all that.

There are a lot of methods that companies are going to be able to help them get through all of this information, including managing the risk of credit if someone wants to borrow money, doing marketing through the database, detecting fraud, and filtering out emails that are spam. It could even be used to help figure out the sentiments and opinions of those who are on your system. All of these, when pulled out and accurate, can help you to gain a good advantage over the competition with your company.

Of course, we have to start somewhere, and this place is with the enormous process of data mining. This data mining is going to help us to break down all of the information in some easy steps that we can follow. To get this to work for our needs, we are going to make it simple and work with five steps that are easy to learn and work with and can help make sure that we get the most out of it.

The first step of this process is that the company is going to go through and collect the data that they need, and then load it up into the warehouse, or another storage area, that they want to use to keep ahold of that information. The type of data that you want to collect is going to be based on what information you would like, and what your overall goal is when you get started with this process as well.

The second step is that we would like to make sure that we are storing and managing the data in the right manner. This is usually going to be done either through the cloud or with some servers that the company is using on their own. Management teams, business analysts, and information technology professionals are then able to

access the data and determine the best method that they can use in order to organize all of that data and learn more information from it.

Then, the company needs to go through and figure out what kind of application software they want to use. There are a number of these that are available for the programmer to choose from, and they can often work with machine learning and the Python coding language to help get the work done. The application software is going to help us to sort out all of the data that we are working on, based on the results from the user.

When all of this is done, the end-user is going to be able to take all of the insights and the information that they have been able to gather up, and then present that data and all of their findings to those who need it. Usually, this needs to be done in a format that is really easy to share, including a table or a graph, so that it is easier for those key people, the ones who really need to use the information, to see what insights are there.

Data mining is going to be a discipline that is able to help us represent a variety of techniques and methods that are used in capabilities that are more analytic than others. This is going to be a useful process because it is able to help us to address a ton of the needs that come in the company, and you can spend some of that time asking different questions and use different levels of human rules or inputs in order to come up with this decision. There are going to be some parts that will come into play when we work with the process of data mining, and a few of these will include:

Descriptive modeling: This is going to be one of the first parts of data mining that we are able to use in our process. When we work with this kind of modeling, we will find that it is responsible for uncovering some of the groupings or similarities that are shared in all of our historical data. This will help us determine the reasons that are behind the failure or the success of a company.

For example, we are able to use this to make it easier to categorize the customers that we are working with. Maybe we will do this by the preference that these customers have to the products that our company is selling or to some of their sentiments overall. There are a variety of techniques that we are able to use to make this happen and that will fit into the idea of descriptive modeling, and some of the ones that we are able to focus on here will include:

- Clustering: This is when we are able to group some of the records that are similar to one another.
- Anomaly detection: This is when we are going to work to identify some of the outliers, and then will determine whether these are important and should be looked at a bit more closely or not.
- Association rule learning: This is when we are going to take some of the records that we have and detect whether there are some kinds of relationships between them and what those relationships are.
- Principal component analysis: This is where we are going to take a look to detect what kind of relationship is going to be present with the variables that we are working on.
- Affinity grouping: This is when we group people together who have a common interest or simple goals at the same time.

Of course, we are able to look at this from another angle as well if we would like. The second part of all this that we are able to focus our attention on is the idea of predictive modeling. This is going to be modeling that is useful because it will go deeper in order to classify some of the events that may happen in the future. It is also a good way to help us estimate some of the outcomes that are not known ahead of time.

A good example of this one would be how we are able to work with credit scoring to make it easier to determine how likely someone is to repay their loan, or if they are more likely to default on that loan ahead of time. The idea that comes with these predictive modeling is that it is going to ensure that we are able to uncover all of the patterns and the various insights that we need in order to make decisions that are better for us. Some of the insights that we are

likely to see here will include the response of a campaign we sent out for marketing, the likelihood of credit default, and even customer churn.

There are going to be a few different techniques that we are able to focus on when it comes to using this option for our needs. Some of the options here for techniques that fit under the term of predictive modeling are going to include:

- Regression: This is going to be a measure of the strength that we see between one dependent variable and a series of variables that are independent.
- Neural networks: This is a complicated type of computer program that can be set up in order to learn, make its own predictions, and even detect some of the patterns that are found in your own data as well.
- Decision trees; These are going to be a type of diagram, that is shaped like a tree, that can help you to make some decisions. The point with this one is that you are able to take a look at each of the branches and see the probable occurrence of each part. This helps us to see which one is the best option for us when making a decision.
- Support vector machine: This is going to be an example of a supervised learning model from machine learning that can help our model to learn the way that it should.

And finally, there is another process that we need to look at a bit more here that is known as prescriptive modeling. Thanks to all of the growth that is going on in our world when it comes to unstructured data and more, and all of the things that we are able to do with this kind of data, the prescriptive modeling option is going to grow a lot in popularity.

Think right now about all of the different sources of data that is unstructured out there that we are able to work with. And these come to us in formats including books, comment fields, web emails, PDFs, audio, and more. This is useful information to work with, but it is

going to need some data mining so that we are able to sort through it and really see what information is inside to help us out.

It is important for us to have a successful method that will help us to parse, filter, and then transform all of this data so that we can really see what is inside of it and add it to our predictive models. It may seem like a lot of extra work along the way, but it is so important to these models because it is going to improve the amount of accuracy that we are going to see with some of the predictions that we make.

In the end, it is so important that we learn how to take a look at data mining, not just as its own separate entity when it comes to data science. It is just as important as some of the others, and it is not a step that we need to miss out on when we are accomplishing some of our work along the way. But it is something that we need to focus our attention on and ensure that we are not going to miss out on if we really want to learn what is going on in our data and how we are able to use the data.

A Look at Data Warehousing

Another topic that we need to spend a bit more time on while we are here is the idea of data warehousing and some of the mining software that we are able to use. You will quickly find that programs for data mining are going to be used quite a bit by a lot of companies, and they are going to be there in order to make it easier for us to analyze the patterns and the relationships that are found in all of the unstructured data that we have. usually, this is going to be done based on the requests that come in from the user. For example, a company is going to be able to use this kind of software to make it easier for them to create classes of information that would help them get things done.

To help us to illustrate this point, let's imagine that there is a restaurant that we want to follow, one who is interested in using some of the processes that come with data mining in order to make it easier to offer the right specials at the right time. They want to make

sure that these specials will hit the right customers and make them the most money possible in the process. This company is going to take a look at all of the information that it has been able to collect and then will create some classes, based on when a customer came to eat, and what they ended up ordering.

Of course, this is just one example of how we are able to use the process of data mining to improve our business. In some other situations, a data miner is able to find a cluster of information based on a logical relationship, or they will look at the associations and sequential patterns in order to draw up a few conclusions about trends that are seen in consumer behavior.

While we are doing this though, we have to remember some of the reasons why we need to focus on warehousing when we do some of our own data mining. Warehousing may sound strange, but it is going to be a simple process where a company is able to pick out one database or type of software in order to centralize all of the raw data that they are focusing on.

With the help of this kind of warehouse to hold the data, the company is able to spin off some of the segments of data that they need over to the right users. These users can then have the right kind of data and use it in analysis, preparing it and getting things ready when it is needed.

However, it is also possible in some situations that we end up in that the analysis is going to start out with some of the data already in place, the data that they want to work with, and then it is possible to work on creating one of these warehouses based on those specifications already from the start. Regardless of how business and some of the other entities out there are going to choose to store and organize their own data, they still want to make sure that they are using it in a manner that is going to support the management decisions later on.

As we can see already, there is a ton of stuff that we are able to do when it comes to working with the data mining process. It may sound like a strange process to work within the beginning and like it is not worth our time, but being able to go through and understand the data that we have and what it all means is going to be so important along the way.

This is going to be the part that shows up in our project of data science that we really need to spend some time focusing on. This is because it will help us to get a better understanding of some of the data that we have and what we can do with it. When we are done with this particular process, we will find that we will have the best possible understanding of that information and how we are able to run it through the algorithm of our choice to get the most out of it all.

Chapter 7: Adding Machine Learning to the Mix



While we are on this topic, we need to spend a bit of time taking a look at something known as machine learning. This is going to be a really important part of the whole process because it allows us to

take all of the information that we have, the information that we have spent a lot of time collecting, organizing, and learning more about, and then puts it through some of the algorithms that will tell us more about the information overall. While all of the steps in this process are going to be really important to help us to gain the insights and patterns that we need for some smart decision making, you will find that this is sometimes considered the fun part, the part that is going to help us to actually put that data to good use.

And we are able to accomplish this with the help of machine learning. This chapter is going to spend some time looking at this buzz word that has taken over the world of business in so many ways. Despite this though, there are a lot of people who have to know the idea of what machine learning is all about or even how they can use machine learning in order to reach some of their own business goals in no time.

To help us get started here, you will find that looking closely at machine learning and what it is all about, and why it is so popular in the world of business today is an important step to get started with. For this one, machine learning is going to be one of the applications of artificial intelligence that will provide our systems with the ability to automatically learn and improve from experience, without us programming it on everything that we want it to be doing in the process. Machine learning is going to focus on the development of computer programs that are going to access data, and then will be able to use this data to help it to learn.

The process of learning in machine learning is going, to begin with, observations, or even data, such as instructions, direct experiences, and examples, in order to look for patterns in data and make better decisions in the future based on the example that we provide. The primary aim is to allow these computers a way to automatically learn without any assistance or intervention from humans, and then you can see that the computer will be able to adjust their actions accordingly to work with this as well.

There are a lot of applications that go with machine learning, and we are going to spend some of our time in this guidebook looking at a lot of the different algorithms and more that you are able to do with machine learning. When you are able to get all of this working together, you will see some amazing results and really see the true potential that comes with machine learning.

There are a lot of different things that you are able to use in machine learning. Any time that you aren't sure how the end result is going to turn up, or you aren't sure what the input of the other person could be, you will find that machine learning can help you get through some of these problems. If you want the computer to be able to go through a long list of options and find patterns or find the right result, then machine learning is going to work the best for you. Some of the other things that machine learning can help out with include:

- Voice recognition
- Facial recognition
- Search engines. The machine learning program is going to start learning from the answers that the individual provides, or the queries, and will start to give better answers near the top as time goes on.
- Recommendations after shopping
- Going through large amounts of data about finances and customers and making accurate predictions about what the company should do to increase profits and happy customers along the way.

Of course, these are just going to be a handful of the times when we would want to rely on machine learning and what it is able to do for us. We can easily see that some of the traditional programs that we are learning how to use as a beginner are going to be too simplistic in order to handle this. They have to work by telling the computer exactly what it should do in a specific situation. And this does work great for a lot of the programs that you want to write. But when you want to add in some artificial intelligence to the mix, you will find that this really is not going to be enough for what you want to do.

In addition, you may find that machine learning is a good thing to use when it is time to handle any of your data analysis, and other parts of a data science project. This part is going to come into play when we need to handle some of the basic and more complex algorithms that are out there in machine learning. There are a lot of algorithms that are going to happen with this one, but knowing how to work with these, and how they fit in with data science and machine learning is going to be important.

Why Should I Use Machine Learning?

As you start to gain some more familiarity with the world of machine learning, you will quickly see that there are a number of benefits of working with the various parts of machine learning. This is probably one of the main reasons that so many companies are interested in machine learning and making it work for their needs. Depending on the questions that you would like to see answered for your company and more, you will find that machine learning can be applied to your business in no time.

To help us out with this, machine learning is going to really help us to simplify some of the steps that come with marketing a product and can even help when it is time to make forecasts on future sales that are more accurate. Machine learning is able to help us out in many ways when it is time to do some of these tasks and more. For example, you will find that going through all of the data that you have collected manually is going to be hard. There is a ton of information in that data for you to explore and learn from, but doing it in a manual way is going to take a lot of time and be really hard to accomplish.

This is where machine learning is going to come into play as well. It is going to give us a look at what insights and patterns are actually in the data, along with some of the past trends that we need to worry about as well. And it is able to do all of this at a much faster rate than we are used to seeing in some of the other parts of our process or by doing it by hand as well.

All of these things are going to come together and ensure that we get some of the results that we would like in no time. you can use this unlimited amount of information to help learn a bit more about your customers and what they would like out of your business, learn the best way to reach them in some of the marketing that you do, what the competition is doing that is so successful and more. Since marketing and all of these other parts are going to be so important in some of the work that you need to do to grow your business, we can already see why machine learning may be the option that you want to go with to get the best results.

In addition, we will find that machine learning is also going to be there to help facilitate some of the predictions and diagnoses that are done in the medical field. This kind of machine learning is a good way for doctors to identify the patients who are at the highest risk for some kinds of diseases, can help us get the best medicine for each case, and so much more.

The way that these programs will work is that they are based on some of the available sets of data on patients, all that is kept anonymous at the time as well, and then it is compared to some of the symptoms that the patient is going through at the time. this is going to help doctors, as well as other medical professionals you work with, add in more precision and efficiency to the job that they are doing. And this is just one of the areas where machine learning is able to help out in the medical field.

Data entry, especially when it is done on a really large set of data, is going to be important but it is going to take too long to accomplish in a manual manner. This is going to be something else that machine learning will be able to step in and help out with. Data duplication and inaccuracy is a big issue that comes up when we do all of this data entry by hand, but automating the process with machine learning can get it done in a quick and efficient manner, and will ensure that it is as accurate as possible.

Another area that we are going to see this machine learning technology do really well is in the financial world. Some of the most common benefits that we will see when we do some research on machine learning in the financial world are going to include fraud detection, algorithmic trading, underwriting of loans and so much more in the process.

In addition, this is the kind of learning that is going to help us with something known as continual data assessments. This is used to help us detect and then analyze some of the anomalies that are going to show up in the financial world, and will help us to improve the precision that we are able to find in our models and other parts as well.

We will also see that machine learning is able to help with detecting spam. This was actually one of the earliest problems that machine learning was able to come in and help with. Spam filters are able to make up new rules, using neural networks, in order to eliminate spam mail and keep your inbox as clean as possible. The neural network is able to learn how to recognize phishing messages as well as other junk mail when it evaluates the rules that are found across an ever-growing network of computers.

The manufacturing industry is even able to benefit from some of the things that we see with machine learning. Manufacturing firms need to have corrective and preventative maintenance practices in place. However, these are going to be inefficient and costly in many cases. This is where machine learning can step in to help, and it is going to be a great tool in creating a highly efficient predictive maintenance plan that keeps the business up and running and doing well. In fact, when the company follows these plans, it is going to minimize the chances of failures that are not expected to happen, which will reduce unnecessary preventive maintenance activities.

Machine learning is also going to help with better customer segmentation and accurate lifetime value prediction. These are going to be some of the biggest challenges that marketers are going to face on a daily basis today. Marketing and sales units are going to have

an enormous amount of data sourced from many channels, but accurate predictions are only going to be found when we look at machine learning.

These are just a few of the benefits that we will be able to work with when it is time to bring out some of the machine learning that we would like to do. And there is so much more that is going to come into play when we use this into the future and figure out the specific ways that we are able to make data science and machine learning go together.

Now that we know a few of the benefits that are available, it is time for us to dive a bit more into machine learning. In specific, we are going to take a look at the three main types of machine learning, and how each of them is meant to work for some of our needs within this kind of field as well. For the most part, we are going to focus on supervised machine learning, which is going to be all about training the models that we have with examples. Then there is also unsupervised machine learning, where we train the algorithm to work by finding the patterns and more that are there all on its own. And finally, we have the reinforcement machine learning that is able to do all of the necessary learning through trial and error.

Supervised Machine Learning



Now that we have had a chance to take a look at machine learning and all that it is going to entail, it is now time for us to really dive into some of this a bit more, and see what we are able to do to make it work for our needs. The first type of machine learning that we are going to spend our time on is known as supervised machine learning. These kinds of algorithms are going to apply some of the information that we have been able to learn about in the past and then will put that knowledge towards some of the new data that we have. This is often done with the help of labeled examples, to help us out with predicting whether something is likely to happen in the future or not.

Starting off with an analysis that happens on a set of data that we know about and uses for training, the algorithm of supervised machine learning that you use here is going to be able to give us a good function to make predictions about any of the values that we give it later on. The system, when we have set it up in the proper manner, is going to be able to provide us with some targets for the new input after you are done with the right amount of training.

From there, the learning algorithm is able to compare the output that it gives with the correct output. And it is also going to be able to find out any of the errors that are there so that it can make some modifications to itself along the way, and provide some better predictions so you know that you can trust it as well.

This is also going to bring up an additional type of machine learning that we are able to work with. This one is going to talk about semi-supervised machine learning. This one is able to combine together what we are talking about with supervised machine learning above, and will add in a bit of the unsupervised machine learning that we will talk about below when we are ready.

For this one to work, it is going to take some of the labeled and unlabeled data and use this in the training that we want to accomplish. in most cases, the labeled data is going to be just a small part of what is being used, and the majority of the data is going to be unlabeled. The reason for this one is that the labeled data, while useful and more efficient, is harder to find and more expensive, and it is often easier to use a combination of the two to help get work done.

The systems that work with supervised and even semi-supervised learning are going to really have more accuracy found in them compared to some of the other options, but there are some limitations that come with it, which is why we have some of the other options to work with as well. This is the type of machine learning that you will want to choose to use when you are able to get ahold of enough labeled data so that the algorithm is easily able to learn as well. Otherwise, the unlabeled data is going to come into play and you will need to focus on semi-supervised or unsupervised machine learning in order to get the work done.

Unsupervised Machine Learning

Now that we have had a chance to take a look at what the supervised machine learning algorithms are able to do, it is time to

take a look at what we are able to do with unsupervised machine learning algorithms. These are going to be the ones that we use any time that the information we have is used to train the algorithm, and it is not going to be labeled or classified. This means that the algorithm, and the system or machine it is on, will need to do the learning on their own, without examples and labeled data to help it make more sense.

Unsupervised learning studies how a system is able to infer a function to describe one of the hidden structures from the unlabeled data. The system doesn't figure out the right output with this one, but it is going to explore the data and then can draw some inferences from the sets of data to describe the hidden structures from the data that is unlabeled.

With this one, we are going to use a lot of data that doesn't have a label on it or any information as to the right answer, and then we are able to send it right through the algorithm and let the system learn along the way. this takes more time and you may end up with some more runs of training and testing before you are done, but it can be one of the best ways to get some strong systems in place to help with your machine learning.

There are a lot of really neat things that we are able to do when it comes to working with unsupervised machine learning. For example, we are able to bring out some algorithms, like the neural networks, that will help us to get things done and learn along the way, without someone having to train the algorithms or teach them all of the steps along the way at all.

Reinforcement Machine Learning

And the final type of machine learning that we are going to spend some time on here is the idea of reinforcement learning. This one is going to be a method of learning that will be able to interact with the environment that is going on around it and then will produce various

actions. This helps the algorithm to discover the errors or rewards that it can get based on these actions.

A good way to compare how this one is going to work is through the idea of trial and error and how we are able to learn from that method. This kind of trial and error is going to add to the search and delayed reward and this learning will be able to make sure that it does what you would like along the way in no time.

When we decide that reinforcement machine learning is the right one for our needs, it is going to be a good one that will allow the machine, as well as the other agents of software that you are using, to act on their own automatically in order to figure out what kind of behavior they should exhibit based on the context that it is in. The goal is that this can happen while maximizing the performance that is seen. It works based on some of the simple reward feedback to the agent so that the agent is able to learn what is best along the way. This is going to be known as the reinforcement signal.

When we take a closer look at this kind of machine learning, there are going to be quite a few similarities that will show up between how the computer learns compared to how a human can learn along the way. This method is set up to help us really work through the process of trial and error, and then the computer is going to be able to use this idea to help them figure out the course of action that will make it the most successful. The more times that it is able to go through this process and be successful, the better it will get with it all and the more accurate the results overall.

As we can see, there are a lot of benefits that are going to show up when we are working with machine learning and all of the things that we are able to do with this kind of learning over time. The more that we want to work with data science and some of the neat things that this process, and the algorithms that are attached to it, are able to do, the more that we will want to focus on machine learning and what this can handle.

There is so much that we can potentially do when it is time to handle some of the machine learning that we want to work with and helping us to figure out the different parts, and how all of them work independently and together, will be a challenge that many data scientists are going to deal with on a regular basis. When it is possible to explore more about machine learning, and some of the different parts that you are able to handle with this language, you will be able to get so much done and really see some of the power that is available through this kind of language as well.

Chapter 8: Completing the Data Analysis



The next step that we need to spend some time on when it comes to working in data science is the idea of the data analysis. This is going to be a fun part to work with because it allows us to learn more about our data and get into some of the different things that we need to know, such as the actual insights and patterns that are in the data.

This is the part where we will actually have a chance to learn a bit about the data. Rather than having to just guess at the data or go through and gather and clean it as we did in the other steps, we now get a chance to send it through the right algorithms and hope that it comes out right on the other side. Of course, there are a number of steps that have to happen to allow us a chance to get accurate results overall, but you will find that working with a data analysis is going to really help us to handle some of the machine learning algorithms we want to use, and learn more about our data than ever before.

With some of this in mind, we need to take some time to explore more about data analysis and what we are able to do with it in our data science project.

What is Data Analysis?



To keep it simple, data analysis is going to be the practice where a company can take their raw data and then order and organize it. When the data is organized in this manner, and run through a predictive model, it is going to help the company extract useful information out of it. The process of organizing and thinking about our data is going to be very important as it is the key to helping us understand what the data does and does not contain at any given time.

Many companies have been collecting data for a long time. They may gather this data from their customers, from surveys, from social media, and many other locations. And while collecting the data is an important step that we need to focus on as well, another thing to consider is what we can do with the data. You can collect all of the data that you would like, but if it just sits in your cloud or a data warehouse and is never mined or used, then it is going to become worthless to you, and you wasted a lot of time and money trying to figure it all out.

This is where data analysis will come in. It is able to take all of that raw data and actually, put it to some good use. It will use various models and algorithms, usually with the help of machine learning and Python, in order to help us to understand what important insights and information are found in our data, and how we are able to utilize these for our own benefit.

You will find that there are a lot of options that we can use when picking out methods to help with data analysis. Often it is not whether the options work, but more about whether they will work on the specific data or the specific problem that we would like to handle along the way. With this in mind, we have to make sure that we carefully choose the right kind of data to go through the information, and we need to make sure that we are not manipulating the data at all.

What we mean with this one is that it is really easy to bring in some of our own thoughts and opinions about the data before we even

start. Sometimes we do this on purpose and other times we may not realize that it is going on at all. But if we are not careful, we will let these manipulations get into the results, and the way that we do things, and then the end results are not going to be as accurate as we would like. Keep all of the biases and the manipulations out, and you will find that the data analysis is going to be better than ever before.

One of the other things that we should really consider when we are doing some of our work here is the quality of the raw data that we are going to use. You will find that the raw data that you choose is going to take on a lot of different forms, and the sources that you use will often depend on your own unique needs and what you are hoping to accomplish from this in the process. You may look on social media posts, focus groups, study groups surveys, websites, and more. These are all great sources, and could potentially help you to get the information that you are looking for.

Once you have been able to gather up that data in the raw form, you will find that even though it is kind of a mess and all over the place, it is still going to be useful for your business to learn and grow. In addition, it is also going to seem overwhelming. This is where the data analysis is going to come into play. It allows us to take some of the workload and some of the intense amount of data, and actually learn from it without feeling overwhelmed or giving up.

When we have all of that raw data in the right form, you will find that this is going to be information that we can use, even though it may not look like it in the beginning. For example, we may find that if we go through and send out a survey to some of our customers, we may get a mess back with lots of answers from people all over the place. But when we have someone go through and sort the answers out, we are able to better see what is going on and can use that information to help us get ahead of the game.

During that unique process of working to get the data as organized as possible, it is likely that you may notice that there are a few big

trends that are emerging as well. These trends are going to be important for us to focus on because they will help us to learn more about which decisions to make, how to work on your business, how to reach your customers, and more. And all of that from some data that may have seemed a bit crazy and all over the place when we first got started.

We can take a look at an example of how this is going to work. When we look at a causal survey about ice cream and what preferences men and women have with this, you may find that more women compared to men had a fondness for eating strawberry ice cream. Depending on what you are doing this whole process for and what results you are hoping to get, it could end up being a major point of interest for the researcher.

Once we know this kind of relationship is in place, we are then able to work with the process of modeling the data. This can be done with a variety of tools, including mathematics, and if you use them in the right way, it is possible that these are going to exaggerate the points of interest so that we can see them a bit easier as we go through this process.

Once we have had some time to go through all of this information and highlight some of the big trends that are found in our data, it is time to work with how it is going to be presented. Of course, we will want to do some kind of textual writeup of the information. This ensures that the people using it to make big decisions are able to see the information and understand what is there. This text needs to include information on the process you took, the resources that you used, and more. This is an important part to add to the process because it ensures that we are going to be able to get the best results and will help whoever is looking at the information to have some background to research as well.

Another thing that we are going to explore a bit more in a later chapter here is that we should add in at least a few visuals to this as well. Options are abundant here and can include things like graphs

and charts. With all of the complex information and relationships that are going to show up in some of the work that we are doing, you may find that you are able to really make some of this work for your needs when you can see it in a graphic format.

If you leave it in the reports and in other documents, it may make sense, but it will take a long time to get through all of that information in order to find out those patterns and insights that you are trying to work with. A much better option to focus on here is going to be to bring in the visuals. These visuals can tell us in just a few minutes the relationships that are found in the documents and in all of that data, something that could take the document a few hours to help explain to us.

This doesn't mean that we need to ignore the reports and do nothing with them. There are plenty of times when these reports are going to be important and we don't want to forget them at all. But we do have to remember that these reports are a good summary of the information, and nothing is able to summarize data better than some good charts or graphs or another visual that works with the data that you have. Have the reports be the backup that helps to explain what is going on with the data, and then have the graphs and charts there to help give us a good idea of the relationships at a glance.

There is so much that has to go on with the data analysis phase of this process, and it is important for us to take some time to learn how this works, and what we are able to do with it. This is basically where we are able to take all of the hard work that we have done in some of the other chapters of this guidebook and put it to use. This is the fun part that includes some of the machine learning, Python, and algorithms to help us see what information is inside.

If we did a good job with some of the other parts of this process and we got this set up in the manner that it should, with high-quality, accurate, and clean information, then we will find that the data analysis phase is going to be a lot easier to work with overall. You will then find that it is easier to pick out the algorithm that you want to

work with, train it and test it, and then put the data through to help you make some of the best business decisions possible.

Steps in a Data Analysis?

Now that we have had some time to talk about what the data analysis is all about and why it is so important to some of the work that we are trying to do within our business, it is time for us to take this a bit further and look closer at some of the steps that you are able to take when we work with this data analysis in the first place. There are actually a number of steps that we will want to take in order to work through this data analysis, and making sure that we get through the right steps, and use them in the right manner, is going to be so critical to ensuring we actually get the right predictions and patterns when we are all done. Some of the steps that we need to follow when working with data analysis will include the following:

- Figuring out what business problem we would like to solve. Whether you have already done a lot of the gathering that is needed of your chosen date, or you are just starting out and you want to know which type of data is the best to gather, you first need to go through and figure out which business problem you would like to solve. This can help to direct the way that this process goes, and ensures that you keep on track with the kind of information that you bring in.
- Searching for the data that we want to work with. Once we have a good idea of the information that we need, and the business problem that we would like to solve, it is time for us to go through and look for the data. There are a number of places where we are able to find this data, such as in surveys, social media, and more, so going out and searching for it here is going to be the best way to gather it up and have it ready to work with on the later steps.
- Cleaning and organizing our data. Since we are usually gathering our data from a lot of different places, and it is often going to come to us in a more unstructured form, it is always a good idea to go through and clean and organize that data. This will make it easier for us to

put that data through the algorithm that we want to use and to make sure that it will all work out.

- Making sure the outliers, missing values, and any duplication information is removed. All of these things are possible to find in your data, and in the beginning, they are not going to seem like they are that big of a deal. But the more of these that are there, the more they are going to mess with some of the results that you are able to get out of your algorithms. We need to spend some time searching through the data to see if those are there, and then make some decisions on how we would be able to handle some of these as well. Taking care of these, rather than just ignoring them, is going to make sure that we are able to get accurate and trustworthy results.
- There are a few methods that we are able to use to make this work for our needs. For example, we may want to spend some of our time working on those outliers and deciding whether or not they are that important. If you look at a little graph about the information in your set of data, and you see there are quite a few outliers that seem to converge together in one general area, this is a good breakthrough to spend some time on. You should at least look into it to figure out if this is something that you should pay attention to and whether it is new information, like a new niche, that is important to you.
- This doesn't mean that all of the outliers that you find are going to be important. Some of them may just be something strange that shows up in the information and if you leave it in there, it is going to end up messing with some of the results that you get along the way. This is why we have to double-check what we are seeing and make sure that it is matching up the way that we think it should, and then decide whether or not those outliers are important for us.
- The next step that we need to spend some time on is creating our own Python algorithm with the help of machine learning. We have already spent some of our time taking a look at machine learning and all of the neat things that you are able to do with this. But in this step, we are going to use machine learning to help us read through the data after it has been trained and tested, and provide us with some of the insights and more that we need.
- There are going to be several algorithms that we are able to choose from when it is time to work with this kind of process, and we

need to know more about our project, and all of the parts that come with it, in order to figure out what is going to be the best for us. Remember that we can't just pick out an algorithm and assume that it is going to work out the way that we want. We have to spend some time training, testing and making sure that the algorithm is going to work the way that we want.

- Look over the patterns and insights that these algorithms are able to find in the data. One of the main points that we will notice when we are handling this kind of data analysis is that it is going to help us take over a large amount of data and then look through it to find all of the important information that is inside of it. The more that we are able to spend our time studying the data, and the stronger the algorithm that we choose, the easier it is to find the insights and all of the hidden and important values that are inside of it. Then we can use all of this information and the patterns that we are given to help us make informed and better decisions overall.
- Create a visualization: This is not a step that you should miss out on at all. These visuals are going to make it easier for those who are in charge of looking over the information to really see the connections and the relationships that show up in that data. This makes it easier for you to really figure out what the data is saying, and to figure out what decisions you should make based on that.
- There are a number of different types of visuals that you are able to work with, but the kind of data that you are sorting through, and the comparisons that you would like to make, are going to help you figure out the right one for your needs. Things like charts, graphs, and more will be the best options for helping you to get through this process as well.

As we can see, there are a lot of different parts that come with our data analysis and what we are able to do with this process. Taking the time to go through and learn more about the data analysis can make a big difference when it comes to how you will run your business. You will be able to use this in order to sort through all of the data you have collected and learned what is hidden inside. These can then be helpful when reaching out to your customers, working on marketing, figuring out how to open up a new niche to explore and

more. But we first need to get ourselves through the process known as data analysis before we are able to use the data that we have been collecting.

As you can imagine from this already, there are going to be quite a few benefits that we are going to be able to see when it is time to work with data analysis. There are many times when you will need to work with this kind of analysis, because it is the part in the data science process that will allow you to finally take some of the data out of your set, and push it through the chosen algorithms, hopefully ones that have been tested and are ready to go

Chapter 9: The Importance of Data Visualizations to Finish the Process



While we are on this process, we need to spend some time taking a look at the importance of data visualizations and how this can help us

to really see some of the results when it comes to our analysis. Sometimes, reading through all of the data that we have, and making it make sense to us is going to be hard when we have to look through a lot of different reports and documents. Sometimes, having this in the picture form, or in some kind of visual, is going to make a world of difference in the success that we have, and the ease of understanding the data that is being presented to you.

This is exactly the spot where we will see the data visualizations come into play. These visuals are going to be helpful because they are able to take all of the data that we were able to collect in the past, and all of the predictions and patterns that the algorithms of Python have been able to show us, and then will place it into a form that is really easy for us to see and understand.

When these visuals are done in the right manner, and usually with the help of a Python library like Matplotlib, we will find that they are a great way to see some of the complex relationships that are present inside of all that data that we have been trying to sort through for such a long time already. For most people, looking through the visuals, rather than just reading through the documents and forms that we have will be a lot easier overall.

The process of data visuals is going to be a presentation of the data so that it turns into a format that we are able to read through and understand better than before. The reason that we are going to work on this one is that it is going to really help us to see things clearly, without guessing or spending hours looking over things, and can help a company make better decisions.

With visuals that add in a bit of interaction with them, we are going to be able to take some of the benefits of this whole concept of visuals even further. What this means is that we are able to not just put a visual out there, but we can add in lots of different types of technologies in order to drill down into the charts and graphs in more details, which will help us to interactively change up the data when we want, such as making changes to the visuals when some more

data is added into the mix. With this all laid out, it is time for us to look closer at the visuals that we have and gain a better understanding of how these are meant to work.

The first thing that we need to take a look at is going to be why these kinds of visuals are going to be so important in the first place. Since the brain is really good at processing out the information that we have in a certain manner, using charts and graphs, and basically any kind of picture or visual in order to help it understand how large amounts of data are going to relate to one another.

Yes, we can read through the information and learn from it. But this process is a whole lot slower than looking at a graph or a chart of this information. And it is likely that we are not going to retain that information as well either. Using the visuals along with some of the reports or the documents that we have is going to be one of the best ways to share that information and learn from it in the process.

Data visualization is a great process to add in with your data analysis. It is a quick and easy way to convey some of the concepts that we have in a universal manner. In addition, we are able to experiment with a lot of different scenarios on the visualization, making a few small adjustments to see what that will do with some of the information that you are working with and the results that you would get from that data.

There are a lot of things that the data visualization is going to be able to help you out with. Some of the ways that we can utilize this data visualization will include:

- It can identify the areas that need the most improvement or your full attention.
- Clarify which factors are going to be the most important when it comes to influencing the behavior of your customer.
- It can help you to take some of your products and understand where you should place them in order to get the most sales.

- It is a great tool to use to make it easier to predict sales volumes throughout different times of the year.

From here, we need to take a look at some of the ways that these data visualizations will be used. You will find that pretty much all industries are going to be able to benefit when it comes to these kinds of visuals. Many of them already know some of the value of collecting this data and that it is important to then analyze the data as well. But now many of them are taking it a bit further and looking at it from a more visual standpoint.

These visual forms are going to make it easier to learn the insights and the predictions that are in all of that data, and that the algorithms were able to find, compared to other options. Some of the ways that these visuals are going to come into play, no matter which industry you are working with, will include:

You can use these kinds of visuals to help make it easier to have a good comprehension of your information and what is found in it, more quickly. By working with a representation of a business that is going to be more graphical in nature, especially when it contains the information that the business has collected it is going to be easier for companies to look through a lot of data in a manner that is much easier and clearer for them to see. And these visuals are also going to be there to make it easier for those who make the big decisions in the company to draw out conclusions from that information.

In addition, it is going to be a lot faster when we are able to analyze information that is found in a more graphical format, compared to analyzing the information that is found in a report or in some kind of spreadsheet. This is in comparison to analyzing information that is found in some kind of report or a spreadsheet. This can also mean that it is easier for these companies to see large amounts of data in a clear and easier manner to understand, and can help the business to address some of the issues that they have, answer questions in a timely manner, and so much more.

One of the other benefits that we are going to be able to see when we work with these data visuals and another way that many companies are going to rely on these visuals is to help us identify some of the patterns and relationships that are going to show up in our data. Even when we are working with extensive amounts of data that seem to be really complicated, we are able to change this around and use the visuals in order to help us make more sense out of that data as well. In fact, when we work with these visuals, we will find that it is a lot easier for a business to handle any of the parameters that are going to be more highly correlated in the process.

One thing that we need to keep in mind with this one is that many times the correlations that we see in these visuals may be obvious, and we can find them and use them for our needs without needing to bring in a visual at all. But then there are going to be times when these correlations are not going to be as obvious, and this is when the visuals are going to be so important to help us figure out these trends. When a company is able to use these options in order to identify the important patterns and relationships, it is going to be a lot easier for them to really focus on the right areas and influence their goals in the best manner possible.

Data visualizations are also able to help a business to discover some trends in the market. Sometimes these trends are going to be just in the market, and other times they can be trends that are specific to the business. When the business is able to find these trends and use them for their advantage, then it is the perfect way to gain an edge over the competition. And this always leads to increasing the bottom line of that company, no matter what industry they are working with.

With the right kind of visualization in place, and with these visuals being used in the proper way, it is much easier to spot some of the outliers that show up in your data. And sometimes, these are the outliers that are going to affect the quality of the product or even the customer churn. You can also use this to address some of the

potential issues that we have before they turn into some really big problems for the company.

In addition, it is possible to use this data visualization to help communicate a story to others. Once a business has been able to uncover some new insights with the help of these visuals, the next step that comes along is to communicate these insights to some others, whether it is the people who make the decisions in the business, the stakeholders, or someone else who is important to this process for your business.

When we are able to work with some of these visuals, whether we are looking at some of the graphs, charts, and other representations of data that are impactful, we know that this kind of work is going to be important when we are working with our data analysis. This is due to the fact that it will help us to keep engagement up some more and will help us to get the information shared and ready to go as quickly as possible.

Before we try to implement some of the new technology that we are looking at and talking about when it comes to data analysis, there are going to be a few important steps that we have to follow in the first place. Not only does a data scientist or a company that is using a data analysis need to have a good grasp on the data they are taking a look at, but they are also going to need to have a good grasp at what kind of data they are looking for, what goals they would like to have, and even their audience.

Being able to prepare the company for the technology that is going to come with these data visuals is going to require that we are able to get a few things in place first, and these will include:

- The data scientist, to start off, is going to need to understand all of the data that they would like to be able to visualize. The most important thing that we are able to understand about the data including how unique the values of data in that column are, which is the cardinality, as well as the size of the data.

- We also need to make sure that we are able to determine what we are working to visualize. We can't just use a random visual and assume that it is going to work with the right option. You can experiment with a few and try them out, but it is always a good option to work with in order to get familiar with the visuals that you are using to keep things organized and safe.
- Another thing that we need to focus on here is that we must know a bit about our audience and have a good idea of how the audience is able to process the information in a manner that is more visual.
- And finally, we need to take a look at using the visual in a manner that is able to take in all of the information and all of the data, and then is able to convey the data in the best option in the easiest form to understand.

After we have had a chance to go through the four parts that are above, you will be able to answer the initial questions that you have about the types of data you want to work with and the audience who is most likely to consume the information. Then we can go through and prepare the data so that it is ready to go straight into the visual that we want to work with.

This is a lot of things to keep in check, and sometimes it is not going to work as well as we would hope. Be ready to take on more work and realize that this is a process that is not going to be as instant and you may have hoped in the beginning. But when it is all said and done, you will have some of the answers that you need. Plus, we will often find that this data can be generated faster than it can be analyzed and processed.

There are factors that we need to consider during this stage as well. This could include something like the cardinality of columns that you are looking to visualize. When we see that the cardinality is higher, this means that there are going to be a large number of unique values. Think about the account numbers at a bank for this one. If we see that the cardinality is low, it could mean that there are a lot of repeat values that show up in the information. This could happen when we work on a column that includes the gender of the person.

There are a lot of benefits that are going to come into play when it is time to work on a data visualization. This is going to be the best way for you to really showcase the information that you have and to make sure that it is in a form that is easy to understand and use all of the time.

You can certainly spend some time just working with just the reports and other options along the way to show off the information and the insights that you are able to get from all of that data and the algorithms that you want to use. But this doesn't help us to see the relationships and more that we need like we can with the visuals. And because there are so many different options that we are able to choose when it comes to the right visuals, you will find that you are certain to find the one that is right for your needs. When you are ready to really work with data analysis or a data science project, and you want to learn how the data is going to relate to one another, then you will need to work through data visualization to help make this happen.

Chapter 10: How Data Science Can Take Businesses into the Future



The final thing that we need to take a look at when it comes to the world of data science is some of the ways that this is able to help us reach the future. There are so many things to consider when it is time to work on our businesses and hope that we can be successful for a long time in the future. In the past, there was a lot of uncertainty that came with the future of a business and how it was going to work. But now we can use data science to help us to see some good results when it comes to our business and learning what decisions to make in the future.

We can quickly see that data science is going to be a great option to use in order to help your business be successful and see some of the benefits that you would like to see in the future. Some of the different things that we are able to see when it comes to the world of data science and all that it is able to do for our needs will include:

Helps You Learn More About Your Customers



One of the top ways that data science is going to be used is as a way to learn more about the customers you have around you. This can work well for marketing purposes, and even to help you figure out new ways to reach your customers and new niches that may be the right one for you. With such a diverse and varied and spread out customer base as most businesses have today, it is important to learn the best way to reach your customers, and exactly what they are looking for when they pick your business or one like you.

All businesses can benefit from learning more about their customers, and your business is not going to be any different along the way. You need to be open to some of the new ideas that are there, and even some of the new demographics that you may not have considered in the past. This is where some of the outliers are going to come into play.

If you see that there are a number of outliers that match up together away from the average, in the same spot, then this is a really good place to look at. It could tell you more about a new product, or even a new niche and demographic, that you should focus on in order to

help grow your business. And if things go right, you may find that it is the right one to use, but also one you and the competition had never thought about before, making it the perfect way to reach your customers where they are, without having a bunch of competition there yet.

Of course, you will find that it can tell us some of the basics about the customers that you are working with as well. This can help us to work with the right customers, reach them right where they are, and ensure that we are going to be able to get to those customers in the way that we need to increase our sales. This is something that has always been a struggle for a lot of businesses, but with the help of some of the work in data analysis and data science, it is now something that we are able to do with a great deal of accuracy compared to the past.

Cut Out Fraud and Other Issues

Many of the financial companies out there are going to use data science in order to help them out in many aspects. Data science and some of the machine learning that goes with it that can cut out on some of the noise and really learn from the past data that is presented to it will be able to help out these industries when it comes to fraud and even helping with underwriting some of the loans out there.

Fraud is a big issue that comes up on a regular basis. And many of those who are going to steal credit card information and more and use it for their own personal gain are getting smarter with some of the options that they decide to use. It is not enough to just set up a maximum amount that the user can spend and say that will catch everything. But having someone watch all of the credit card transactions that are going through for a company would be overwhelming and not all that effective.

You will find though that with the help of some of the machine learning algorithms, we are able to feed in information about past

transactions that we know are fraudulent and allow the algorithm to learn. The algorithm is then able to go through and pick out, from some of the new transactions that show up, which options are going to be fraudulent, which ones need to be checked by a human, and which ones are going to be safe. This can help save credit cards, banks, and other financial institutions billions a year.

Another way that many of these companies are going to start using data science in the future is to help them when picking out who to give loans to or not. This may seem a bit unfair and like it takes some of the personal touches out of the whole thing when a computer gets to decide. But these can help to cut down on the bad applications that are accepted and reduce the amount of waste that the bank or financial institution is taking on overall.

For the customer, this is going to result in some faster processing times. If you are accepted, you will find out much faster, and usually with a better interest rate, than what was happening in the past. So, while this may seem like something that is only going to benefit the lender, it is actually something that can benefit both parties.

Learn How to Cut Out Waste

All companies have some kind of waste. But it isn't always what you may think. Just because you are not throwing away countless boxes of paper or trash on a daily basis does not mean that you are not dealing with a lot of waste that is cutting into your profits and making you less competitive because you have to charge more of your products.

There could be a waste in how long a process takes. If you can find a faster method to use to create a product or service, one that keeps the same quality that the customer has come to expect, and does not cause harm to the employees in the process, then this is something that you need to consider looking into. And that is where data science can help.

Do you notice that there seems to be a lot of downtimes, or waiting time in between the different steps of the process? Or maybe there is a kind of bottleneck that is going to come up when we work on this process, and while everyone else is speeding up, one place is stopping it all and can't keep up, and the others after it are just waiting around? This is not an efficient use of your employees and machine to just have them standing around, and data science will be able to help you figure this out and make some of the changes that are necessary.

What about when one of the machines in your company ends up breaking down? This can often happen right in the middle of the day and then all operations have to stop and no employee can work, even though they are getting paid until it is fixed. Machine learning and data science are able to come into this and make predictions on when a part in the machine is going to break down, based on how often you use it and the age of the part. Then you can schedule the best time to fix that part, rather than trying to get it done at the worst times.

These are just a few of the places where we are likely to see a lot of waste when it comes to our business, no matter what kind of industry we are in. It is important to learn how to reduce this kind of waste, and more, as much as possible, so that we are able to see some good results in the process. When we reduce the waste, we reduce the cost that we experience, and this can make it a whole lot easier to keep prices down, remain competitive, and really see some of the results that we want in the process.

How to Handle Your Competition

In our modern world, there is always a lot of competition. In the past, the competition that you had to work against was usually those that were nearby. You didn't have to compete against anyone who was more than a short drive away because people would not go that far to get what they wanted. Now, with modern technology and the use of the internet, your competition could be on the other side of the world.

The ease of using the internet to set up your own shop is going to cause some of the problems as well. So many small businesses and at home businesses will start up in your industry as well, which means more types of competition that you have to fight against. And all of this is going to make it really hard to see some of the success that you would like unless you can find ways to distinguish yourself from all of the noise.

This is where the data science project is going to come in. You can use this to figure out some new ways to market to your customers and get them to choose you over someone else. You can use this in order to learn more about the other competitors and what they are doing that you could do better, or at least in a different manner, to beat them out. You can use data science to help you figure out a new niche to reach, help you to figure out which products you would like to sell, and so much more. Sometimes even just learning more about the industry that you are in can be a good goal when we are handling some of the work of data science.

Make Better Business Decisions



Many companies are choosing to go with data science, and all of the other steps that we have spent some time on in this guidebook, in order to help them make some smarter business decisions overall. In the past, there was always an element of risk that came with making some good business decisions. If you wanted to be able to pick out the next product that should sell or figure out which direction you had to take for your business, it was hard. There had to be a good mix of business smarts, knowledge of being in the industry for a long time, and a ton of risk to get it all to work.

Data science is able to change this. As long as you are able to use the process of data science in the proper manner, you will find that it is going to help you to reduce your risks, while also helping to make decisions that you are confident are the right ones for your needs. You will have all of the data, and a well-trained algorithm, to help you get this done, and will ensure that you are going to be able to make smart business decisions overall.

No longer will you need to worry about the amount of risk that comes with some of the decisions that you have to make. No longer will you

need to focus on the pros and the cons of something. All of this is going to be done with the right project in data science, and with the help of good data and strong machine learning algorithms along the way. When you can combine these together, you will be able to get the assurance that you need that you are always making the best business decisions all the time.

Data science is really changing up every industry that it touches. Whether you are working in retail, the medical field, manufacturing, marketing, financial fields, or somewhere else, you will find that data science is able to help you do your work better. Your customers are going to notice, the competition is going to see that you are blowing right past them, and it won't be long before you are able to really see some improvements.

Right now, there is just a lot of speculation out there about how this data science is going to work and how far it can take us in the future. But seeing what it has been able to do so far, and how much it has been able to help out a lot of businesses overall, can really give us some hope for how great this is going to be in the future as well.

Conclusion

Thank you for reading *Python Data Science*. I hope that you enjoyed this book and found out the information that you need about how to work with Python along with your data science project today.

In this guidebook, we spent a lot of time taking a look at the process of data science, and why it has become so important to so many different businesses over time. It is a great process, one that allows a company to take in as much information as they can, and then learn some of the valuable insights and predictions that they need to help them run their business better.

As we discussed a bit more in this guidebook, the data that we are able to gather is not going to be as structured and easy to sort

through as it was in the past. This is going to make it a bit harder for us to go through and actually sort through the data manually as we could in the past. It also means that because the data does not need to be organized to be available, we are able to find a lot more of this information overall as well.

But with all of this data and all of the organization that is missing from the data, it is a small wonder that we are able to learn anything out of the data at all. But this is exactly where data science is going to come into play and will help us to get started in no time. When we work with data science and all that it is going to provide to us along the way, we will learn that it is possible, through a series of steps like gathering the data, cleaning it, and sending it through the right algorithms, that will help us to learn more about what is found in the data and what we are able to do to make it work for our needs.

This guidebook took some time to explain more about what the data science process is all about. We explored data science and some of the basics of this language. We then worked with the process of Python and how this can work with the data science project to get the best results. From there we moved on to some of the benefits of data science, how to work with data mining, the importance of data cleaning and organizing and so much more.

This is not where the process is going to stop though. We also get to spend some time looking at the basics of machine learning, and how this ties into one particular step of data science, the data analysis. This is the fun step, where we get to work on training some of our models to make them work the way that we want and to ensure that we actually learn the right predictions and more from that data. We spent a good deal of time looking at these topics, along with the data analysis to make sure your information was ready to go.

The end of this guidebook will take look at what the importance of data visuals is all about, and some of the ways that we are likely to use data science and the machine learning process to help us get started on the right track. All of this is going to come together to help

us create some of the best codes that we need to finally get the right results.

There are many books on this topic, and we hope this one provides you with the information and skills that you need to get the best results. If you found this guidebook helpful to you, make sure to leave a review!

Table of Contents

Introduction

Chapter 1: The Basics of Data Science

Why Does My Business Need Data Science?

A Look at Data Science

Data Science and Business Intelligence

The Lifecycle of Data Science

Chapter 2: How Does the Python Language Fit In With Data Science?

Chapter 3: The Best Python Libraries to Help with Data Science

NumPy

SciPy

Pandas

Matplotlib

Scikit-Learn

Theano

TensorFlow

Keras

Chapter 4: Gathering Your Data

Know Your Biggest Business Problem

Places to Look for the Data

Where to Store the Data?

Chapter 5: Organizing and Cleaning the Data

What is Data Preparation?

Why Is Data Preparation So Important?

Steps Involved for Data Preparation

Chapter 6: A Look at Data Mining

The Process of Data Mining

A Look at Data Warehousing

Chapter 7: Adding Machine Learning to the Mix

Why Should I Use Machine Learning?

Supervised Machine Learning

Unsupervised Machine Learning

Reinforcement Machine Learning

Chapter 8: Completing the Data Analysis

What is Data Analysis?

Steps in a Data Analysis?

Chapter 9: The Importance of Data Visualizations to Finish the Process

Helps You Learn More About Your Customers

Cut Out Fraud and Other Issues

Learn How to Cut Out Waste

How to Handle Your Competition

Make Better Business Decisions

Conclusion