# https://lh4.googleusercontent.com/YAkJn7VqqmAS8Vv8-JktR7RmdchyEDaNQUZSWUg3c1E1rKm9PSRoZwMP1pBLNrGNZ_sdrzO83F1Kg_RJHQis9STLyEw8TN12v56np9ZB7Qmwb_WAsfmGdxKFLcSFMUnIe-5ZJczX **BOOK TITLE: Machine Learning in Biotechnology using Python.**

***Subtitle:***

# **AUTHOR: Saleh Alkhalifa**



# **ABOUT THE AUTHOR**

In order for us to write your author bio we need a few details. Please remember that your answers should be ***relevant*** to the book. Your bio helps sell the book so please *only* include *relevant* information:

|  |  |
| --- | --- |
| **Full name** | Saleh Alkhalifa |
| **What is your job title?** | Data Scientist, Team Manager |
| **What is the name of the organization you work for?** |  |
| **What is your skillset (as relevant to the book)?** | Data Scientist & Software Engineer |
| **In which industry do you work?** | Biotechnology, Technology |
|  |  |
| **What University degree do you have?** | BS, MS |
| **What professional qualifications do you have?** | BS, MS |
| **Please detail your *relevant* work history** | Data Scientist |
|  | Senior Associate Data Scientist |
|  | Associate Scientist |
| ***Relevant* projects you have worked on:** |  |
| 1 Applications of Machine Learning in Biotechnology | 5 Applications of Deep Learning in Forecasting |
| 2 Applications of NLP in Resume Development | Please see my [GitHub](https://github.com/alkhalifas) for my public-facing list. |
| 3 Anomaly Detection in Instrument Failures |  |
| 4 Prediction of Biological Potency in Small Molecules |  |
| **Would you like to include your social media details (optional)?** |  |

# **Can you recommend a good technical reviewer for your book?**

[INSERT HERE: *Name & Contact details*]

# PART ONE: BACKGROUND RESEARCH

# **TARGET AUDIENCE**

Describe your target reader: what you assume about their knowledge of the topic, related topics, and technical topics generally; why they want to learn the technology; what will they want to do with it?

***Answer the following****:*

Who is your audience?

|  |  |
| --- | --- |
| 1 | There are two main audiences for this book which consist of students and professionals studying and working in the scientific space. Students in various scientific fields such as chemistry, biochemistry, biology and biotechnology are one of the main audiences targeted here. There currently exists a large student population within a crowded scientific field prompting many to learn additional individualized skillsets in order to be more competitive in the postgraduate market. Most of these students have defaulted towards artificial intelligence as the current AI market has shown little to no signs of contracting. The second main target population are scientific professionals that are already established within the pharmaceutical and biotechnology sectors. Many scientists have expressed strong desires to learning and apply AI methods within their roles to drive efficiency and agility. |
| 2 | We can assume that the audience are within the scientific field, either from an academic or industry setting. We can also assume that the audience have an interest in having some technical background in data science in conjunction with their areas of work. |

What is important to them?

|  |  |
| --- | --- |
| 1 | The main issue the audience is dealing with is finding a comprehensive single-source of materials to educate them in data science as it relates to the scientific field. There are many platforms that teach python, and many websites that teach machine learning, however, there are currently no platforms that give readers an introduction to python and machine learning from the biotech perspective. |
| 2 | The audiences from this book tend to get challenged on “getting started”. This book would remedy this problem by offering a comprehensive and up-to-date guide on getting started in a new field that is not yet that well documented. |

# **COMPETITIVE BOOK TITLES**

What is unique about your book? You will need to look on Amazon at books that have been well-received – what are the top three market leading books that your book will compete with? Examine the description, table of contents and book reviews.

**List the books here**:

|  |  |
| --- | --- |
| 1 | [INSERT HERE: Competitor 1] |
| 2 | [INSERT HERE: Competitor 2] |
| 3 | [INSERT HERE: Competitor 3] |

# 

Please ensure that you have looked at the **description**, **table of contents** and **book reviews** for each of these books.

PART TWO: BOOK OVERVIEW

# **OVERVIEW**

The long description is the device we use to describe the book on Amazon. Writing it is fairly systematic. Please answer the following questions using only one sentence.

|  |  |  |
| --- | --- | --- |
| **TEMPLATE** | **EXAMPLE:** Hands-On Blockchain Development with Hyperledger | **Your turn...** |
| Explain / Introduce the tech | Hyperledger Fabric and Hyperledger Composer enable organizations to create private, permissioned blockchain networks. | Applications of machine learning within the biotechnology and scientific space. The main focus is to have a single-source of material to prepare a prospective data scientist operating in biotech with all the tools and templates to drive success in their careers and companies. |
| Why would a developer want to learn it? |  | The Biotechnology field is one of the most prosperous in the world, and has been changing quite drastically over the last few years. Pricing pressures from the government and other institutions are driving drug prices down, which has caused a major surge in company interests when it comes to savings. Savings are currently accomplished through data-driven decisions (machine learning), however many scientists are not able to understand / execute on these techniques. In order for scientist to differentiate themselves from others and contribute to the major ‘cost savings’ endeavors companies have embarked on, this book will bridge the gap in both the technical skillset as well the general mentality a data scientist must have within the field. |
| Why should they buy this book? | Developers working with Blockchain will be able to put their knowledge to work with this practical guide. The book provides a hands-on approach to implementation and associated methodologies that will have you up-and-running, and productive in no time. | Scientists and developers working with machine learning in biotechnology will be able to put their knowledge to work with this practical guide. The book provides a hands-on approach to implementation and associated methodologies that will have you up-and-running, and productive in no time. |
| Product approach | Complete with step-by-step explanations of essential concepts, practical examples and self-assessment questions, you will begin by exploring the blockchain evolution, including an overview of relevant blockchain technologies | Complete with step-by-step tutorials, explanations of essential concepts, relevant examples directions on future projects and approaches to similar problems. |
| Product Breakdown: In 2 sentences, describe the “journey” the book takes the reader on. Look at your section headings for help | You'll learn how to configure Hyperledger Fabric on a cloud platform, understand the architectural components of Hyperledger Fabric, and how they are configured to build private blockchain networks, and the applications that connect to them. You'll then build a network and application from scratch, and learn how to implement smart contracts in chaincode. | You will learn how to develop, tune and deploy sophisticated machine learning, and deep learning models from scratch for the purposes of automation and cost savings as they relate to the biotechnology space. You’ll then deploy these models to the cloud through various frameworks such as Heroku and AWS. |
| By the end of this book you will... | By the end of this book, you will be able to build and deploy your own decentralized applications using Hyperledger, addressing the key pain points encountered in the blockchain life cycle. | By the end of this book you will be able to build and deploy your own machine learning models to automate tasks and make predictions using Heroku and AWS. |
| Anything else you would like to add? |  |  |

# 

# **LEARNING OUTCOME - WHAT WILL THE READER LEARN AND DO?**

Consider the competing books; in particular the **description**, **table of contents** and **book reviews**. Decide what the key learning objectives will be for your book. List them below:

|  |  |
| --- | --- |
| 1 | Get stared with Python3 and SQL |
| 2 | Understand the meanings and differences between various areas of machine learning |
| 3 | Be able to develop a machine learning predictive model from scratch using Python |
| 4 | Be able to fine-tune models to optimize their performances for various tasks |
| 5 | Be able to deploy, evaluate and monitor a model in the cloud. |

PART THREE: BOOK STRUCTURE

Using your **overview**, and **learning outcomes** now decide on the structure of your book? What are your start and end points?

# **GENERAL STRUCTURE**

**Divide the book into approximately 3 parts**. The **learning outcomes** you listed previously will help to inform these. These “parts” are a group of chapters that work toward the same goal. Each part will consist of 3-5 chapters. For example: A book on Building Machine Learning Systems with Python might be split into 5 parts as follows: “The Basics”; “Book Learning”; “Numbers, Forecasts and Recommendations”; “Sound and Vision” and finally, “Practical Matters”.

**WRITE YOUR PART HEADINGS BELOW:**

|  |  |
| --- | --- |
| 1 | Getting Started with Data |
| 2 | Developing and Training Models |
| 3 | Deploying Models to Users |

# **CHAPTER OUTLINE**

**Each chapter should have a clear focus**. Each chapter title should clearly state what aspect of the overall topic the chapter deals with.. Continuing the example of *Building Machine Learning Systems with Python* your section on “Book Learning” might be broken down into 4 chapters as follows: “Clustering – sorting text into groups”, “Topic Modeling – creating non-exclusive groups”; “Logistic Regression – evaluating text quality”; “Bayes Classification – sentiment analysis”. *PLEASE NOTE: Chapter titles appear on Amazon*

|  |  |  |  |
| --- | --- | --- | --- |
| **Part 1: Getting Started with Data** | | **Page Count** | **Cum Sum** |
| **1** | **Introduction** | 2 | 2 |
| **2** | **Python and the Command Line** | 8 | 10 |
|  | *General Coding Tutorial* | 10 | 20 |
| **3** | **Intro to SQL** | 4 | 24 |
|  | *General Coding Tutorial* | 8 | 32 |
| **4** | **Visualizing Data** | 3 | 35 |
|  | *Biotech Coding Tutorial* | 13 | 48 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Part 2: Developing and Training Models** | | **Page Count** | **Cum Sum** |
| **5** | **Introduction to Machine Learning** | 5 | 53 |
| **6** | **Unsupervised Machine Learning** | 6 | 59 |
|  | *Biotech Coding Tutorial* | 11 | 70 |
|  | *Biotech Coding Tutorial* | 8 | 78 |
| **7** | **Supervised Machine Learning** | 20 | 98 |
|  | *Biotech Coding Tutorial* | 9 | 107 |
|  | *Biotech Coding Tutorial* | 10 | 117 |
| **8** | **Natural Language Processing** | 11 | 128 |
|  | *Biotech Coding Tutorial* | 10 | 138 |
|  | *Biotech Coding Tutorial* | 10 | 148 |
| **9** | **Deep Learning in Biotech** | 7 | 155 |
|  | *Biotech Coding Tutorial* | 5 | 160 |
| **10** | **End-to-End Model Development** | 2 | 162 |
|  | *Biotech Coding Tutorial* | 15 | 177 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Part 3: Deploying Models to Users** | | **Page Count** | **Cum Sum** |
| **11** | Deploying with Flask | 7 | 184 |
|  | *Flask Coding Tutorial* | 5 | 189 |
| **12** | Deploying to the Cloud | 15 | 204 |

**PLEASE ADD OTHER PARTS/CHAPTERS ONLY IF YOUR TOPIC REQUIRES IT**

PART FOUR: DETAILED OUTLINE

# PART 1: Getting Started with Data

This chapter will describe the basics of Python, SQL, and translating raw data into meaningful visualizations and representations as the first step of a strong data science project. The vast majority of readers and users are generally overwhelmed with the vast data science content found on the internet or in print; this book remedies this issue by focusing in on the most important and valuable

### **CHAPTER 1:** Introduction - 2 pages

### 

### DESCRIPTION: The objective of this chapter is to briefly introduce the reader to the objectives of this book, the main topics we will touch on, and a sense of how the book is structured to best orient them with the content.

### Level: Basic

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: Getting Started
2. HEADING 2: Setting Up
3. HEADING 3: Things to Know

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Understand the scope and audience of this book
2. SKILL 2: What tools will they need to be successful with this book
3. SKILL 3: Understand the notation that this book will use

### 

### **CHAPTER 2:** Python and the Command Line - 18 pages

### 

### DESCRIPTION:

In this chapter, readers will be introduced to the most important take-away examples and teachings of the Python programming language. Users will be given a quick crash course on data types, iteration, and basic programming. In addition, a basic introduction to the command line will be provided.

### Level: Basic

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: Objects
2. HEADING 2: Data Types
3. HEADING 3: Module 1- Data & Python
4. HEADING 4: The Command Line

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: A brief introduction to the concept of objects in Python
2. SKILL 2: An introduction to datatypes in python
3. SKILL 3: An introduction to important Python operations and best practices
4. SKILL 4: Introduce the command line and how best to use it in conjunction with Python code

### 

### **CHAPTER 3:** Introduction to SQL - 12 pages

### DESCRIPTION:

This chapter will focus on the use of SQL, a fundamental querying language allowing data scientists to retrieve large blocks of data from enterprise-level databases. In this section, we will touch on the basics of SQL, types of databases, as well as a hands-on module showing a real-life biotechnology example.

### Level: Basic

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: What is SQL
2. HEADING 2: Relational and Non-Relational Databases
3. HEADING 3: Module – Introduction to SQL

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will learn the main idea behind SQL databases
2. SKILL 2: Readers will get a sense of the different types of commonly used databases
3. SKILL 3: Introduce the basics of SQL and the most commonly used commands

### 

### **CHAPTER 4:** Visualizing Data - 16 pages

### 

### DESCRIPTION:

The objective behind this chapter is to give users a holistic overview of the different methods of visualizations. In this chapter, we will cover simple plots, heat maps, timeseries plots, Sankey diagrams, as well as small/large molecule visualizations.

### Level: Basic

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: The Six Steps of Visualization
2. HEADING 2: Visualization Libraries to Know
3. HEADING 3: Module – Data Representation and Visualizations

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will get a better idea of the steps needed to develop a strong visual component
2. SKILL 2: Readers will learn about the most common libraries they are encouraged to explore
3. SKILL 3: Readers will be able to apply the code and visualizations to their own data

### 

# PART 2: Developing and Training Models

In this section, readers will be introduced to the process of parsing data, and training models. We begin with an introduction to the two forms of machine learning, and reinforce the definitions a number of modules/tutorials using real world examples in the biotechnology industry.

### **CHAPTER 5:** Introduction to Machine Learning - 5 pages

### 

### DESCRIPTION:

The objective behind this chapter is to give a brief and concise introduction to the concepts of machine learning, how it fits into the larger AI picture, and the two main types of machine learning: Supervise and Unsupervised.

### Level: Intermediate

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: What is Machine Learning
2. HEADING 2: Common Machine Learning Libraries
3. HEADING 3: Supervised vs. Unsupervised ML
4. HEADING 4: The Machine Learning Roadmap

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will understand the main concepts behind machine learning
2. SKILL 2: Readers will know the most common libraries and when to use them
3. SKILL 3: Readers will understand when to use SML, and when to use UML
4. SKILL 4: Readers will be able to determine when to use which algorithm

### 

### **CHAPTER 6:** Unsupervised Machine Learning - 25 pages

### 

### DESCRIPTION:

This chapter describes the area of unsupervised machine learning. A few concepts will be introduced, some common terminology discussed, and finally, two modules/tutorials will be presented. Readers will be introduced to a historical unsupervised application through a famous dataset, followed by a novel dataset based on a real-world example.

### Level: Intermediate

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: What is Unsupervised Machine Learning
2. HEADING 2: Transformations
3. HEADING 3: Module – Dimensionality Reduction for Drug Toxicity
4. HEADING 4: Clustering
5. HEADING 5: Module – Breast Cancer Prediction via Clustering

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will walk away with the main concepts and ideas behind unsupervised learning
2. SKILL 2: Readers will learn and apply transformation techniques
3. SKILL 3: Readers will learn and apply dimensionality reduction techniques to real-world data
4. SKILL 4: Readers will learn and apply clustering techniques
5. SKILL 5: Readers will learn and apply clustering techniques to real-world data

### **CHAPTER 7:** Supervised Machine Learning - 39 pages

### DESCRIPTION:

This chapter describes the area of supervised machine learning in which readers will be introduced to some common terminology, an understanding of model fitting, regression and classification techniques. This chapter will also encompass two modules/tutorials – one pertaining to regression and another to classification.

### Level: Intermediate

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: What is Supervised Machine Learning
2. HEADING 2: Overfitting & Underfitting
3. HEADING 3: Regression
4. HEADING 4: Module – Applications of Regression for Property prediction
5. HEADING 5: Classification
6. HEADING 6: Module – Breast Cancer Prediction via Classification

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will walk away with the main concepts and ideas behind supervised learning
2. SKILL 2: Readers will learn and apply transformation techniques
3. SKILL 3: Readers will understand regression models and when to use them
4. SKILL 4: Readers will be introduced to a number of regression models or a real-world example
5. SKILL 5: Readers will understand classification models and when to use them
6. SKILL 6: Readers will be introduced to a number of classification models or a real-world example

### **CHAPTER 8:** Natural Language Processing - 31 pages

### 

### DESCRIPTION:

In this chapter, the reader will be given a deep dive into applications of natural language processing to parse documents, retrieve insights, and generate topics and word clouds. Through the use of the NLTK and Spacy libraries, readers will be able to apply a number of these techniques to their respective tasks.

### Level: Intermediate

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: What is Natural Language Processing?
2. HEADING 2: Generating Insights from Text
3. HEADING 3: Common NLP Applications
4. HEADING 4: Module – Sentiment Analysis on Drug Side Effect Ratings
5. HEADING 5: Module – PENDING

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will be able to understand NLP and its associated terminology
2. SKILL 2: Readers will learn to generate insights using their own data
3. SKILL 3: Readers will understand some of the most common NLP applications
4. SKILL 4: Readers will be introduced to an exciting real-world application of NLP and sentiment
5. SKILL 5: PENDING

### **CHAPTER 9:** Deep Learning - 20 pages

### DESCRIPTION:

Within this chapter, readers will be introduced to the concepts and applications of deep learning as it relates to the biotechnology industry. Within this chapter, emphasis will be placed on some of the main concepts, deep learning components, and the general types of neural networks. Readers will then be walked through a module/tutorial concerning chemical property predictions using deep learning!

### Level: Advanced

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: What is deep learning?
2. HEADING 2: Activation Functions
3. HEADING 3: Neural Network Architectures
4. HEADING 4: Neural Network Parameters
5. HEADING 5: Module – Applications of Deep Learning for Chemical Property Prediction

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will be given a strong understanding of deep learning
2. SKILL 2: Readers will learn about activation functions and when to use them
3. SKILL 3: Readers will learn about the different architecture styles
4. SKILL 4: Readers will be introduced to the main DL parameters and how to use them
5. SKILL 5: Readers will be able to see a deep learning model, and apply it to their own cases

### **CHAPTER 10:** Timeseries Forecasting - 20 pages

### DESCRIPTION:

Within this chapter, readers will be introduced to the concepts and applications of time series forecasting as it relates to the biotechnology industry.

### Level: Advanced

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: What is timeseries forecasting?
2. HEADING 2: Simple Exponential Smoothing (SES)
3. HEADING 3: Autoregressive Integrated Moving Average (ARIMA)
4. HEADING 4: Long Short-Term Memory Models (LSTM)
5. HEADING 5: Module – Applications of Timeseries Forecasting in Biotechnology

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will be given a strong understanding of timeseries forecasting
2. SKILL 2: Readers will learn SES and how to apply in with their own data
3. SKILL 3: Readers will learn ARIMA and how to apply in with their own data
4. SKILL 4: Readers will learn LSTM and how to apply in with their own data
5. SKILL 5: Readers will be able to see a full implementation of a forecasting model in a real-world example

# PART 3: Deploying Models to Users

Thus far we have discussed Python, using data and developing models; within this chapter we will explore how these models can be moved to production and made available to an end-user. We will explore four commonly used platforms to deploy a machine learning model: AWS, Heroku, GCP, and Python Anywhere.

### 

### **CHAPTER 11:** Deploying Models with Flask - 15 pages

### 

### DESCRIPTION:

Within this chapter we will cover the basic steps to deploying a machine learning model using the Flask framework. Readers will be introduced to the concepts of APIs and WebApps, and proceed to develop a full application.

### Level: Intermediate

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: What is Flask?
2. HEADING 2: APIs and WebApps
3. HEADING 3: Module – Creating a Flask Application

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will understand how the flask framework operates
2. SKILL 2: Readers will understand when to use an API vs a WebApp
3. SKILL 3: Readers will learn to develop a full flask application from scratch

### 

### **CHAPTER 12:** Deploying to the Cloud - 20 pages

### 

### DESCRIPTION:

Within this final chapter, readers will be guided through a series of tutorials on deploying their machine learning models using Flask within a number of cloud-based providers: AWS, GCP, and Heroku.

### Level: Intermediate

### Main Chapter Headings (3-5 main chapter headings)

1. HEADING 1: Deploying to the Cloud
2. HEADING 2: Deploying with AWS
3. HEADING 3: Deploying with GCP
4. HEADING 4: Deploying with Heroku

### Skills learned: *For each heading, insert what the reader will learn to DO in this chapter?*

1. SKILL 1: Readers will learn the general process of ‘deploying’ to the cloud
2. SKILL 2: Readers will be guided on how to deploy a model using AWS
3. SKILL 3: Readers will be guided on how to deploy a model using GCP
4. SKILL 4: Readers will be guided on how to deploy a model using Heroku

### 