Skills needed to be a Data Scientist/Machine Learning Specialist

A person wearing glasses and smiling at the camera

Description automatically generated

Introduction

*Harvard Business Review* suggested that data scientist is the “sexiest job” of the 21st century. Questions such as “What makes someone a data scientist?” and “What skills are needed to be a data scientist?” started appearing. Through this, I hope to bring some light on what a data scientist is using Stack Overflow survey data from 2019. More specifically, we are going to see what programming languages, platforms, web frames, database are used most as well as the future projected use.

The dataset I got from Stack Overflow contained data from people with different occupations. One of the first things I did was filter the dataset to only include those whose job is data scientist or machine learning specialist. I then explored the data more and decided to analyze several fields: programming languages, platforms, databases, and web frames. This will hopefully give a sense on what skills are needed to be a data scientist or a machine learning specialist.

Part 1- What are the most used and desired to use programming languages?

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Programming languages used now Programming languages desired next year

We can see here that **JavaScript, HTML/CSS, SQL and Python** are the most used programming languages for both 2019 and what is forecasted for 2020.

Part 2- What are the most used and desired to use platforms?

A picture containing fence

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Platforms used now Platforms desired next year

We can see here that **Linux and Windows** are the two most used platforms in 2019 with several others following behind. However, we see a new trend in desired used of platforms for 2020. We see that Docker has switched place with Windows and that AWS and Android are following closely behind.

Part 3- What are the most used and desired to use web frames?

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Web frames used now Web frames desired next year

We can see here that **jQuery** is the most used web frame in 2019. However, **React.js** is the most desired web frame to use for the year 2020. However, there are a few other web frames that are also used now and desired to use in 2020.

Part 4- What are the most used and desired to use databases?

A screenshot of a cell phone

Description automatically generatedA picture containing fence

Description automatically generated

Databases used now Databases desired next year

We can see here that **MySQL** is the most used database in 2019 with others trailing pretty far behind. However, we can see how the playing field levels for what is desired the next year. We see how **PostgreSQL, MySQL, MongoDB** have pretty equal shares for what is desired.

Conclusion

In this article, we took a look into Stack Overflow’s 1029 survey data to see the several fa tors that makes up a data scientist or a machine learning specialist.

1. The most popular programming languages are **JavaScript, HTML/CSS, SQL and Python.**
2. **Linux and Windows** are the two most used platforms in 2019.
3. **jQuery** is the most used web frame in 2019. However, **React.js** is the most desired web frame to use for the year 2020.
4. **MySQL** is the most used database in 2019 with **PostgreSQL, MySQL, MongoDB** rising up and having pretty equal shares for what is desired in 2020.

I hope that this article sheds some light on what skills you have to work on to be a data scientist or a machine learning specialist.

References

Stack Overflow Survey data: <https://insights.stackoverflow.com/survey>

Github: