Vol 06 / Issue 01 / March 2024 / pp 56 – 70
ISSN : 2676-234X / EISSN : 2716-9006

Social Empowerment Journal
University Amar Telidji of Laghouat-Algeria

The Modern Methods of Data Analysis in Social Research: Python Programming Language and its Pandas Library as an Example- a Theoretic Study

Ikram Harouni
University of Algiers 2 -Algeria, e-mail: ikram.harouni@univ-alger2.dz
Labratory of Prevention and Ergonomics, University of Algiers 2

ORCID: 0009-0002-0863-5484

Received: 09/01/2024; **Accepted**: 21/03/2024, **Published:** 31/03/2024

Keywords

Data analysis; Pandas library; Python; Social research.

Abstract

This study aims to highlight the significance and various dimensions of utilizing the Python language in the field of social sciences. Python is widely recognized as one of the most frequently employed programming languages, primarily due to its user-friendly nature and suitability for data analysis in social research. In this study, particular emphasis is placed on the Pandas library, which holds a prominent position in data analysis for social research. The library offers robust tools for data analysis, manipulation, and the execution of mathematical and statistical operations in an efficient and accessible manner. The study has concluded that learning programming languages and using them in data analysis in social sciences has become crucial nowadays to understand, analyze and interpret the contemporary society, as well as to generate new scientific perspectives and produce knowledge that corresponds to the rapid changes in the social field.

Corresponding Author : Ikram Harouni
DOI 10.34118/sej.v6i1.3806

مجلة التمكين الاجتماعي
جامعة عمار ثليجي الأغواط – الجزائر

المجلد 06 / العدد 01 / مارس 2024 / ص ص 56 – 70
ISSN : 2676-234X / EISSN : 2716-9006

الأساليب الحديثة في تحليل البيانات في البحوث الاجتماعية
لغة Python ومكتبة Pandas نموذجا

إكرام هاروني

جامعة الجزائر2 - الجزائر، البريد الإلكتروني: ikram.harouni@univ-alger2.dz

مخبر الوقاية والارغونوميا، جامعة الجزائر2

ORCID: 0009-0002-0863-5484

تاريخ الاستلام: 2024/01/09 - تاريخ القبول: 2024/03/21 - تاريخ النشر: 2024/03/31

الكلمات المفتاحية

تحليل البيانات؛
مكتبة بانداس؛
لغة بايثون؛
البحوث الاجتماعية.

الملخص

سنحاول من خلال هذه الورقة البحثية تسليط الضوء على أهمية وأبعاد استخدامات لغة 'بايثون' في العلوم الاجتماعية، كأحد لغات البرمجة الأكثر استخداما، باعتبارها لغة سهلة التعلم، قوية، وسهلة التطبيق، لتحليل البيانات في البحوث الاجتماعية، والتركيز على مكتبة Pandas باعتبارها واحدة من المكتبات الأكثر استخداماً في تحليل البيانات في تلك البحوث، حيث توفر وسائل قوية لتحليل وتحويل البيانات وإجراء العمليات الحسابية والإحصائية على البيانات بشكل سهل وفعال. وقد توصلت الدراسة إلى أن تعلُّم لغات البرمجة وتوظيفها في تحليل البيانات في العلوم الاجتماعية أصبح أمرًا حاسمًا في وقتنا الحاضر، لفهم وتحليل وتفسير المجتمع المعاصر ولتوليد رؤى علمية جديدة وإنتاج معرفة توافق التغيرات المتسارعة في المجال الاجتماعي.

المؤلف المرسل: إكرام هاروني

DOI 10.34118/sej.v6i1.3806

1- Introduction:

The art of divination and interpreting trends through data analysis dates back to ancient times. The ancient Egyptians used census data to improve tax collection efficiency and accurately predict annual Nile floods. Since then, practitioners of data science have created a unique and distinct field in which they work, known as data science.

With the evolution of computer technologies and the explosion of information, data analysis methods in social research have significantly evolved. Data has become big data,fast and complex, which has necessitated the development of new software and tools to meet these challenges and process them quickly. It has become essential for social science researchers to master some of these computer tools and algorithms.

More generally, with the availability of social networks, computing has made it possible to calculate and model the social as if the collected data offered better access to the real social than any survey, census, or poll (Boullier, 2015) . This has led to the establishment of massive computer infrastructure in many countries to represent archives, data processing and publication, as well as the use of new data sources outside of traditional censuses and surveys. Digital change is presented as an inference framework and allows social sciences to understand the changes that occur in societies, (Boullier, 2015). analyze heterogeneous effects, and arrive at results and recommendations that help understand social behavior and public opinion, analyze social and political trends, and develop public policies and social solutions.

Thus, the digital age has imposed new challenges on social science researchers, forcing them to use new tools to analyze large amounts of data from online sources. This evolution is often referred to as a shift towards "computational social science (Damien, 2019).

Python programming language is one of the most used languages in social sciences, as big data analysis with Python has become an essential element for social science researchers. Among the characteristics of this language, it is powerful and easy to use, and it stands out for its vast and diverse libraries dedicated to data analysis In particular, the 'Pandas' library offers powerful tools for data analysis (Brooker, 2019) and transformation, as well as for performing mathematical and statistical operations on data in an easy and efficient way.

In this study, we will try to highlight the importance and dimensions of the uses of this language and its use in data analysis in social sciences, with a focus on the Pandas library. Thus, we pose the following question:

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

What is the importance and dimensions of the uses of the Python language in data analysis in social sciences?

2- The concept of programming language and its importance in social sciences:

Programming is one of the essential tools in the world of technology. It is the means by which humans communicate and interact with computers and control them. A programming language comprises a set of rules, commands, and instructions used to create computer programs. It is used to convert concepts and ideas into understandable and executable instructions by the computer.

Programming languages differ in their structure, system, and method of writing and execution. They can be classified into two types:

- Low-level languages: These are languages used to control electronic devices and interact with hardware.

- High-level languages: These are languages that are easier to use and are widely used for the construction of programs and applications.

The difference between programming languages, such as Java, Python, C#, and C++, lies in the fact that they offer different functionalities, tools, and supporting libraries, and are used in different fields. Programmers choose the language that best fits their research needs and level of experience.

Learning a programming language is a valuable skill for social science researchers in the era of information technology, for several reasons:

Programming significantly enhances social scientists' ability to effectively analyze sizable internet-derived data through automation, data handling and visualization generation using statistical tools like R and Python.

*- **Reproducibility**: Programming allows researchers to document and share their analysis workflows, permitting others to reproduce studies and validate findings, ensuring transparency and accountability in science.*

*- **Flexibility**: Programming allows researchers to customize their data analysis workflows to address specific research questions. They can use existing packages or libraries, or write their own code to perform specialized analyses.*

*- **Collaboration**: Programming encourages collaboration among researchers by allowing them to share code and work together on projects. This can lead to more efficient and productive research outcomes.*

Programming can also help visualize data in social science research in several ways, enabling them to present their results more effectively:

*- **Automation**: By using programming, researchers can automate the process of creating data visualizations, allowing them to create charts quickly and efficiently. This can save them time and help them explore their data more effectively (Echout, 2023).*

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

- ***Customization***: *Programming enables researchers to customize their visualizations to better illustrate their results. They can adjust colors, fonts, labels, and other elements to create clear and attractive visualizations that fit their research questions.*

- ***Interactivity***: *Programming allows researchers to create interactive visualizations, allowing users to explore data in more detail. This can be particularly useful for visualizing complex or multidimensional data, where users can manipulate the visualization to uncover patterns and relationships that may not be apparent in a static image.*

Python is a high-level programming language that was created in 1991 by 'Guido van Rossum', and it has continued to evolve since then. Python is distinguished by its ease of use and learning, as well as its wide range of libraries and tools that facilitate data analysis and visualization. It has become extremely popular in the fields of social sciences due to its user-friendliness and versatility (programmez, 2021).

An example of how to analyze survey data using Python and the Pandas library in the context of social research:

Suppose we conducted a survey on the attitudes of young adults towards environmental conservation. The survey collected data on various demographic variables such as age, gender, and education, as well as responses to questions about environmental beliefs and behaviors. We want to analyze this data to answer some research questions such as:

- *What is the distribution of environmental beliefs among young adults?*
- *What factors are associated with environmental behaviors?*
- *How do environmental beliefs and behaviors vary by demographic variables?*
- *To answer these questions, we can use Python and the Pandas library. Here's an example code:*

Figure N° 1

An example of how to analyze survey data using Python and the Pandas library in the context of social research

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

```
import pandas as pd

# load the survey data
survey_data = pd.read_csv('survey_data.csv')

# calculate the distribution of environmental beliefs
belief_distribution = survey_data['belief'].value_count
s(normalize=True)

# calculate the association between demographic variabl
es and environmental behaviors
behavior_association = survey_data.groupby(['age', 'gen
der', 'education']).mean()['behavior']

# calculate the variation in environmental beliefs and
behaviors by demographic variables
belief_by_demographics = survey_data.groupby(['age', 'g
ender', 'education']).mean()['belief']
behavior_by_demographics = survey_data.groupby(['age',
'gender', 'education']).mean()['behavior']
```

```
# print the results
print('Environmental belief distribution:\n', belief_di
stribution)
print('\nAssociation between demographic variables and
environmental behaviors:\n', behavior_association)
print('\nEnvironmental beliefs by demographic variable
s:\n', belief_by_demographics)
print('\nEnvironmental behaviors by demographic variabl
es:\n', behavior_by_demographics)
```

Source: *(Python SourceCode).*

In this code, we used the read_csv() function of Pandas to load the survey data from a CSV file. Then we used different Pandas functions to perform the necessary calculations. For example, we used the value_counts() function to calculate the distribution of environmental beliefs, the groupby() function to calculate the association between demographic variables and environmental behaviors, and the groupby() function again to calculate the variation in environmental beliefs and behaviors by demographic variables.

Finally, we printed the results using the print() function. With this code, we were able to analyze the survey data and answer the research questions effectively and accurately.

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

Since its creation, the Python language has undergone significant evolution, with the release of many different versions, as well as the development of additional libraries and tools that make programming and data analysis easier. In addition, Python is now available on various operating systems and devices, making it accessible to everyone for use in different fields.

Since Python is open source, developers and social science researchers can benefit from open source software developed using Python, modify it, and develop it to meet their specific needs.

The main steps in data analysis using Python in social sciences include data cleaning, exploration, modeling, and visualization (moez, 2023) :

Data Collection → Data Cleaning → Data Exploration → Data Modeling → Evaluation → Interpretation → Communication

Python offers libraries that can support each of these steps, making it a powerful tool for data analysis.

Researchers can use several libraries provided by this language, such as:

- ***Pandas** : for tabular data manipulation and analysis.*
- ***NumPy**: for numerical calculations and array operations.*
- ***Matplotlib**: for creating plots and visualizations.*
- ***SciPy**: for scientific calculations and statistics.*
- ***Scikit-learn**: for machine learning and data exploration.*

Python provides social scientists a flexible environment for data analysis through versatile libraries offering diverse tools and features. While certain libraries support tasks like natural language processing and social network analysis, additional techniques may apply depending on research needs. Overall, Python presents a powerful platform for social science data interpretation leveraging comprehensive libraries and strong community support.

3- Limitations of using Python in social sciences:

Social science researchers primarily focus on writing relatively simple programs (scripts) that execute a series of instructions (such as variable recoding, merging datasets, text document analysis, etc.) to organize their data into a model that they can analyze. They typically focus on learning to use the available tools effectively rather than developing new tools.

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

For example, suppose a social science researcher wants to study the impact of age, gender, and income on technology mistrust. The research steps could be as follows:

- *Collect data from a random sample of individuals, including age, gender, income, and level of technology mistrust.*
- *Store the data in a CSV file.*
- *Use Python and the Pandas library by social science researchers to read the file and convert it into a dataframe.*
- *Recode variables to make them analyzable, for example, converting gender to 0 and 1, where 0 represents females and 1 represents males.*
- *Merge datasets to group data related to age, gender, income, and technology mistrust level into a single data frame.*
- *Analyze the data using statistical and mathematical analysis tools such as regression tests and factorial tests to determine the relationships between variables and evaluate the relative impact of each variable on technology mistrust.*

Most experimental social science relies on tabular data with variables in columns and observations in rows. However, Python lacks a built-in tabular structure, confusing some social scientists. They thus utilize the Pandas library, which offers tabs, column renaming and merging different data types - mirroring capabilities in Stata and R crucial for social research handling tabular information. (Arne, 2023).

In addition, Pandas also serves as a gateway to many other tools that are likely to be used by social scientists, such as visualization libraries (seaborn and ggplot2) and the statsmodels library for econometrics.

Here's an example:

Suppose we have a CSV file containing employee data, including their name, age, salary, and email address, but with no tabular structure or order. Here's how Pandas can be used to convert this data into a tabular format.

Figure N° 2

An example how Pandas can be used to convert this data into a tabular format

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

```python
import pandas as pd
import numpy as np

# read in data from a CSV file
data = pd.read_csv('data_file.csv')

# calculate mean, median, and standard deviation of a v
ariable
mean = np.mean(data['variable'])
median = np.median(data['variable'])
std_dev = np.std(data['variable'])

# create a histogram of a variable
import matplotlib.pyplot as plt
plt.hist(data['variable'])
plt.show()
```

***Source**: (Python SourceCode).*

In this example, pd.read_csv() is used to read the data from an unstructured CSV file. Then, data.columns is used to rename the table columns with more meaningful names. data.head() can be used to display the first few rows of the data in tabular format.

By using the Pandas library, social scientists can leverage the power and flexibility of Python to efficiently process and analyze tabular data. The Pandas library enables researchers to work with data more efficiently and easily, thanks to the strong support it receives from the NumPy library. NumPy allows it to perform algebraic and statistical operations quickly and efficiently. It also uses the data structure defined in NumPy (i.e., rows and columns). (Takefuji, 2021).

In general, here are some features and functionalities of the Pandas library:

- *The library provides a data structure called Data Frame, which makes it easy to import and manipulate data.*
- *It offers the necessary capabilities to import data from different file formats and load them into memory.*
- *It facilitates data preprocessing operations such as data cleaning, handling missing values, and performing exploratory operations on the data.*
- *It allows for reshaping the data structure.*
- *It facilitates merging or splitting data into multiple Data Frames.*

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

- *It offers the ability to convert data to different formats, handle missing data, and provide flexible methods for transformation, filtering, aggregation, statistical analysis, and other various operations.*
- *It manages and analyzes independent and large-scale data.*

Some of the features offered by Pandas for large-scale data analysis include:

- *Progressive loading of data: the chunksize option in the read_csv() function allows for progressively loading the data rather than loading it all into memory. This makes it easy to handle large-scale data without exhausting resources.*
- *Performance enhancements: Pandas has been significantly optimized to improve its - performance for large-scale data analysis, using specialized tools such as Dask and Apache Arrow.*
- *Compression techniques: Pandas supports data compression techniques such as vertical and horizontal compression, which help reduce data size and speed up analysis operations.*

In addition, Pandas can be used in parallel with other technologies such as Spark, Hadoop, etc., to extend its capabilities for large-scale data analysis.

Social science researchers can also turn to other libraries to load, analyze, and visualize data in the social sciences field. These can also be used for analyzing and understanding social phenomena, studying social relationships, and developing machine learning models in the social sciences.

4- The dimensions of using Python in data analysis in social sciences:

4-1- The use of Python in analyzing big and complex data in social sciences:

The use of Python in analyzing complex and voluminous data in social sciences is considered a powerful and effective tool. It allows for efficient work on this type of data and precise analysis, including various types of analysis such as:

4-1-1- The application of Python in textual data analysis in social sciences:

Content analysis is one of the most important research techniques in social sciences. Content analysts consider data not as representations of physical events, but as texts, images, and expressions created to be visible, readable, interpretable, and therefore processed according to their meaning. Therefore, they must be analyzed based on this usage. (klauws, 2004, p. 10)

Content analysis differs from other methods of inquiry by analyzing texts in their context of use. In academic fields such as anthropology, linguistics, social psychology, sociology of knowledge, and the smaller field of communication studies, many practical activities have emerged: therapy, advertising, politics, arts, etc.

Rapid technological changes and new ways of accessing content have impacted how analysts examine and interpret information. To keep pace, they have developed

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

novel methodologies aligned with these shifts, enabling accurate and thorough analysis and sound conclusions. Computer systems that simulate coordinated social processes aid this approach. Vast archives of press, polls, reports, records and texts are now interlinked online networks that can be evaluated from various perspectives, as social systems themselves have moved onto computers.

This evolution requires a redefinition of content analysis, taking into account content - the research objective and understanding contemporary society through its texts.

These scientific methods will be used together to analyze heterogeneous effects and arrive at results and recommendations that help understand social behavior and public opinion, analyze social and political trends, and develop public policies and social solutions.

For example, in a 2016 study on internet use in the coverage of US presidential elections in digital media, researchers analyzed content over a period of time in digital media sources such as news sites. The study showed that there were clear differences in how US presidential elections were covered in digital media. Researchers were able to identify the main issues that media coverage focused on, as well as psychological analysis of candidates, parties, and political issues. (JEFFREY GOTTFRIED, 2016).

4-1-2- The applications of Python in analyzing temporal and spatial data in social sciences:

Temporal data in social sciences refer to data that evolve over time, such as historical data and temporal statistics on social, political, and economic transformations, such as changes in social, economic, or political behaviors. Spatial data in social sciences refer to data related to geographical locations, areas, and spaces, such as data on migration, geographic distribution of populations, local development, violence, etc. (Cornelia, et al.)

Temporal and spatial data in social sciences are used to analyze the relationships between events and social phenomena over time and space, as well as to provide a better understanding of the mutual relationships between different social forces and to identify the geographical characteristics in which different social forces may have an impact.

The Python language provides a number of tools that can be used for the analysis of temporal and spatial data. The most common ones are:

- ***Pandas**: a powerful library for temporal and spatial data analysis in Python. Pandas provides functions for conversion, sorting, processing, and statistical analysis of temporal and spatial data.*

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

- ***NumPy**: a useful library for mathematical and statistical operations in Python, providing functions for statistical analysis and processing of temporal and spatial data.*
- ***Matplotlib**: a library for creating graphics in Python, providing functions for creating different graphics for temporal and spatial data.*
- ***Seaborn**: a library for creating advanced graphics in Python, providing functions for creating advanced graphics for temporal and spatial data.*
- ***GeoPandas**: a library for geographic data analysis in Python, providing functions for geographic and spatial data analysis.*
- ***PySAL**: a library for spatial data analysis in Python, providing functions for geographic and spatial data analysis.*
- ***Statsmodels**: a library for data analysis in Python, providing functions for statistical analysis, forecasting, and statistical modeling of temporal and spatial data.*

Python has been used to effectively analyze geospatial and temporal social science data, like a study identifying high crime areas in Karachi through Python and GeoPandas analysis of spatial data. This guided officials in understanding crime locations, timing and priorities for prevention by enabling rapid and accurate identification of hotspots from the geographical information. (mehran khan, 2019).

The main steps of the study included collecting geospatial data on violent crime from different sources and storing them in a spatial database, and then using Python and GeoPandas to analyze the data and identify hotspots and areas with the highest rates of violent crime. This step was done using geostatistical techniques integrated into GeoPandas.

4-1-3- The use of Python in statistical modeling and forecasting in social sciences:

Statistical modeling and forecasting in social sciences involve using statistical and mathematical tools to analyze social data and create models that show relationships between different social variables. These models are used to predict human behavior and social interactions, such as crime, and to provide guidance to individuals and organizations in evidence-based decision making. (Fatima Dakalbab a, 2022, pp. 202-222)

Statistical modeling in social sciences includes regression analysis and multiple regression analysis, analysis of independent variables, factor analysis, multiple factor analysis, and other statistical methods. These statistical tools are applied to quantitative and qualitative data collected on different social variables, such as age, gender, income, education, culture, beliefs, values, and behaviors.

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

Python is useful for statistical modeling and forecasting in the social sciences given its precision, speed and libraries like Pandas that enable efficient analysis of temporal, spatial, multi-dimensional data and tests like regression while noting accurate results also depend on quality data, modeling methods and analyst expertise beyond the programming language alone.

4-1-4- Social network analysis using Python:

Social networks are modern online tools for communication and social interaction used for various purposes. They allow individuals to create profiles, connect, communicate by posting and commenting. Over time, social networks have become integral to many people's lives worldwide.

Specialized social network analysis studies how ideas spread online by examining relationships, connections, interactions, websites and communication patterns between individuals and groups through various techniques and large data analysis to understand trending online behaviors applicable to fields like media, marketing, education and social sciences.

For example, the study conducted by' Jonah Peretti 'and his colleagues to improve the lmeme tracker"aimed to understand the l'nformation cycle"between blogs and mainstream media, focusing on political campaigns that follow the spread of memes on the web. Memes have different characteristics of bending and deviation as they spread in different contexts. Peretti's work involves using different types of effects that memes can leave and treating them as strings that can be found all over the web. This approach has been used to produce temporal measures, day by day, or even minute by minute with Twitter (Boullier, 2015).

4-1-5- Using Python in geographic and spatial data analysis in social sciences:

Python is widely used in geographic and spatial data analysis in social sciences, and their visualization on maps, using different libraries that support this type of analysis. Python can be used in geographic and spatial data analysis in social sciences in several ways, including: (Avipsa & Fouché, 2019)

- ***Use of libraries such as GeoPandas, Shapely, and others**: which help in the analysis of spatial and geographic data, creation of spatial models, and analysis of relationships between spatial variables.*
- ***Design and implementation of distance analysis**: Python can be used to design and implement distance analysis, using libraries such as PySAL and others, which help in the analysis of spatial relationships between variables, and provide graphs and graphical visualizations of the results.*
- ***Visualization of spatial and geographic data on maps**: Python can be used to visualize spatial and geographic data on maps, using libraries such as Folium,*

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

Geoplot, and others, which help in creating interactive maps and visualizing data in a visual and intuitive way.

- ***Analysis of geographic data using networks**: Python can analyze geographic data using networks through libraries like NetworkX and OSMnx. These helps integrate geographic analysis, build complex models, and analyze relationships. Python is also used in social sciences, for example analyzing crime data and its spread in locations. With growing IoT, real-time crime data can be collected and geographically analyzed using Python libraries to build smarter, safer cities by understanding crime patterns.*

Combining crime data with social media insights provides more information about risky areas, aiding community safety. Python libraries like Matplotlib and Folium facilitate clear visualization of results, resulting in accessible and expandable data frameworks. These enable analysis by officials and police of neighborhoods' crime levels and associating characteristics, further investigating what may contribute to criminal acts in certain locations.

4-1-6- Using artificial intelligence and machine learning based on Python in social science research:

Artificial intelligence and machine learning based on Python can be used in social sciences to perform many tasks, including: (Fatima Dakalbab a, 2022).

Big Data Analysis: Python and AI/ML techniques can be used to analyze large amounts of data used in social sciences, such as social network analysis, social interactions, and social forecasting.

- ***Prediction of Social Behavior**: AI/ML techniques can be used to predict social behavior, such as identity authentication, psychological analysis, and prediction of social situations.*
- ***Text Analysis**: Python and AI/ML techniques can be used for text analysis, such as sentiment analysis, text classification, and keyword recognition.*
- ***Classification and Prediction**: AI/ML techniques can be used for data classification and prediction of outcomes, such as user classification, survey response prediction, and social behavior prediction.*
- ***Social Network Analysis**: AI/ML techniques can be used for social network social scientists can strongly benefit from Python .Machine learning helps analyze large surveys, identifying variable relationships and grouping respondents. AI assists by detecting and addressing missing data to enhance analysis rigor. Overall, artificial intelligence and machine learning empower*

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

social scientists to reveal useful insights from surveys more productively than traditional statistical methods alone. Researchers can also:

1- Understand semantics: semantic technology

Semantics is an important concept in understanding unstructured data, as it refers to a branch of linguistics and logic that deals with meaning on social media, people use symbols like emojis to show how they feel without words. Programs understand different types of posts by looking at what's in them. Analyzing social media uses AI to find patterns in the data. Experts then interpret the patterns by linking language knowledge with computer results. Combining meaning, machines, and social studies helps get more from online information

2-Predict emotions:

Text processing converts raw text into a usable form using techniques such as encoding and filtering. For example: The analysis will first process the text and then use a tool like NLTK Vader or a trained model to examine sentiment (python-social-media-analytics, 2017).

5- Results:

Based on our analysis, it is evident that learning programming languages and utilizing them for data analysis in social sciences have become crucial in contemporary times. This is necessary to comprehend, analyze, and interpret the complexities of modern society, as well as to generate novel scientific perspectives and produce knowledge that aligns with the rapid changes occurring in the social domain.

Furthermore, given the exponential growth of digital environments and the consequent generation of vast amounts of data that cannot be effectively analyzed using traditional computing methods, it is imperative for social science researchers to incorporate these data and employ digital tools for analysis and interpretation. Python, with its ease of learning and extensive features, offers a multitude of powerful and user-friendly tools and libraries that facilitate the analysis of large datasets, creation of intricate statistical models, forecasting, spatial and geographic analysis, and intuitive visualization of spatial data on maps.

6- Conclusion:

In conclusion, it is expected that the reliance on programming languages for data analysis in social sciences will continue to grow in the future. This trend will enable a deeper understanding and more precise interpretation of social phenomena. With the emergence of the metaverse and the increasing integration of artificial intelligence, researchers will increasingly rely on data analysis techniques based on AI, machine learning, and deep neural networks. Such techniques will empower them to collect and

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

analyze vast amounts of data, extracting crucial insights for a more accurate and effective understanding of social phenomena.

Across all social fields, including education, health, economics, politics, and society, data analysis will play a pivotal role in enhancing social, health, and economic services and improving government policies and social planning. As a result, data analysis will contribute significantly to enhancing the overall quality of life, promoting social and economic progress, and achieving sustainable development worldwide.

7- Referencias:

Arne, G. (2023). Python for Social Scientists. Consulté le july 2023, sur real python: https://realpython.com/python-for-social-scientists /

Avipsa, R., & Fouché, E. (2019, november 5-8). In-Database Geospatial Analytics using Python, ARIC'19.

Boullier, D. (2015). the social sciences and the traces of big data:society,opinion, or vibration? revue francaise de science politique, 65(5-6), 805-828.

Brooker, P. D. (2019). Programming with Python for social scientists. Récupéré sur sage: https://us.sagepub.com/en-us/nam/programming-with-python-for-social-scientists/book259581

Cornelia, T., Anna Laura, R., Linda, H., Nina, B., Thierbach, Anna Laura, R., & Linda, H. C. (s.d.). Spatial Analysis in the Social Sciences and Humanities. Towards Integrating Qualitative, Quantitative and Cartographic Approaches. Historical Social Research, 39(148), pp. 09-22-.

Damien, T. (2019). Doing Computational Social Science with Python: An Introduction. Consulté le July 2023, sur SSRN: https://ssrn.com/abstract=2737682

Echout, M. (2023). Automating with python: a developer's guide for 2023, phython, carrer advice. Récupéré sur https://www.learnenough.com/blog/automating-with-python

Fatima Dakalbab a, M. A. (2022). Artificial intelligence and crime prediction: A systematic literature review. Social Sciences & Humanities Open, 6(1), pp. 202-220.

JEFFREY GOTTFRIED, M. B. (2016). The 2016 Presidential Campaign – a News Event That's Hard to Miss. Consulté le July 31, 2023, sur pew research center: https://www.pewresearch.org/journalism/2016/02/04/the-2016-presidential-campaign-a-news-event-thats-hard-to-miss/

klauws, k. (2004). content analysis : an introduction to its methodology (éd. 2). new delhi, india: sage inc, london.

mehran khan, r. a. (2019). Hotspot Analysis of Crimes Using GIS; A Case Study of District Abbottabad. Récupéré sur https://www.researchgate.net/publication/330690348_Hotspot_Analysis_of_Crimes_Using_GIS_A_Case_Study_of_District_Abbottabad

moez, A. (Éd.). (2023). NLTK Sentiment Analysis Tutorial for Beginners. Consulté le july 2023, sur datacamp: NLTK Sentiment Analysis Tutorial for Beginners

programmez. (2021). Consulté le juillet 2023, sur programmez le magazine des developpeurs,Bientôt un langage Python 2 fois plus rapide ?: https://www.programmez.com/actualites/guido-van-rossum

python-social-media-analytics. (2017). Python Social Media Analytics. Consulté le 2023, sur document.pub: https://dokumen.pub/python-social-media-analytics-9781787121485-1787121488.html

Takefuji, Y. (2021). SCORECOVID: A Python Package Index for scoring the individual policies against COVID-19. Healthcare Analytics, pp. 100-105.

***Citation**: I. Harouni. **The Modern Methods of Data Analysis in Social Research: Python Programming Language and its Pandas Library as an Example- a Theoretic Study**. Social Empowerment Journal. 2024; 6(1): pp. 56-70. https://doi.org/ DOI 10.34118/sej.v6i1.3806*

***Publisher's Note:** SEJ stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.*

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en