Introduction to Python for Accounting Applications

Min-Yuh Day, Ph.D,
Associate Professor

Institute of Information Management, National Taipei University

https://web.ntpu.edu.tw/~myday
Min-Yuh Day, Ph.D.

Associate Professor, Information Management, NTPU

Visiting Scholar, IIS, Academia Sinica

Ph.D., Information Management, NTU

Director, Intelligent Financial Innovation Technology, IFIT Lab, IM, NTPU

Associate Director, Fintech and Green Finance Center, NTPU

Artificial Intelligence, Financial Technology, Big Data Analytics, Data Mining and Text Mining, Electronic Commerce
Course Syllabus
National Taipei University
Academic Year 112, 1st Semester (Fall 2023)

• Course Title: Python for Accounting Applications
• Instructor: Min-Yuh Day
• Course Class: ACC2, NTPU (3 Credits, Elective)
• Details
  • EMI Course
    (3 Credits, Elective, One Semester) (U2004)
• Time & Place: Wed. 6, 7, 8, 14:10-17:00(B3F17)
• Google Meet: https://meet.google.com/ofh-iosa-ehd
Course Objectives

1. Understand the fundamental concepts of Python for Accounting Applications.
2. Equip with Hands-on practices of Python for Accounting Applications.
Course Outline

• This course introduces the fundamental concepts and hands-on practices of Python for Accounting Applications.

• Topics include
  • Introduction to Python for Accounting Applications,
  • Python Programming and Data Science,
  • Foundations of Python Programming,
  • Data Structures, Control Logic and Loops,
  • Functions and Modules,
  • Files and Exception Handling,
  • Data Analytics and Visualization with Python,
  • Obtaining Data From the Web with Python,
  • Statistical Analysis with Python, Machine Learning with Python,
  • Text Analytics with Python with LLMs,
  • Applications of Accounting Data Analytics with Python, and
  • Applications of ESG Data Analytics with Python.
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Subject/Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2023/09/13</td>
<td>Introduction to Python for Accounting Applications</td>
</tr>
<tr>
<td>2</td>
<td>2023/09/20</td>
<td>Python Programming and Data Science</td>
</tr>
<tr>
<td>3</td>
<td>2023/09/27</td>
<td>Foundations of Python Programming</td>
</tr>
<tr>
<td>4</td>
<td>2023/10/04</td>
<td>Data Structures</td>
</tr>
<tr>
<td>5</td>
<td>2023/10/11</td>
<td>Control Logic and Loops</td>
</tr>
<tr>
<td>6</td>
<td>2023/10/18</td>
<td>Functions and Modules</td>
</tr>
<tr>
<td>7</td>
<td>2023/10/25</td>
<td>Files and Exception Handling</td>
</tr>
<tr>
<td>8</td>
<td>2023/11/01</td>
<td>Midterm Project Report</td>
</tr>
</tbody>
</table>
# Syllabus

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Subject/Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>2023/11/08</td>
<td>Data Analytics and Visualization with Python</td>
</tr>
<tr>
<td>10</td>
<td>2023/11/15</td>
<td>Obtaining Data From the Web with Python</td>
</tr>
<tr>
<td>11</td>
<td>2023/11/22</td>
<td>Statistical Analysis with Python</td>
</tr>
<tr>
<td>12</td>
<td>2023/11/29</td>
<td>Machine Learning with Python</td>
</tr>
<tr>
<td>13</td>
<td>2023/12/06</td>
<td>Text Analytics with Python and Large Language Models (LLMs)</td>
</tr>
<tr>
<td>14</td>
<td>2023/12/13</td>
<td>Applications of Accounting Data Analytics with Python</td>
</tr>
<tr>
<td>15</td>
<td>2023/12/20</td>
<td>Applications of ESG Data Analytics with Python</td>
</tr>
<tr>
<td>16</td>
<td>2023/12/27</td>
<td>Final Project Report</td>
</tr>
</tbody>
</table>
Teaching Methods and Activities

• Lecture
• Discussion
• Practicum
Evaluation Methods

• Individual Presentation 30 %
• Group Presentation 30 %
• Case Report 20 %
• Class Participation 10 %
• Assignment 10 %
Required Texts

• Allen B. Downey (2016), Think Python: How to Think Like a Computer Scientist, 2nd Edition, O’Reilly Media
Reference Books

1. Frederick Kaefer and Paul Kaefer (2020), Introduction to Python Programming for Business and Social Science Applications, SAGE Publications


Other References

• Python, https://www.python.org/
Green Finance and Sustainable Finance
Evolution of Sustainable Finance Research

SDGs: Sustainable Development Goals

AI for Environmental, Social, and Governance (AI4ESG)

Source: Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle Belgrave et al. (2020) "AI for social good: unlocking the opportunity for positive impact." Nature Communications 11, no. 1: 1-6.
AI for Social Good (AI4SG)

Sustainability

SDGs

CSR

ESG
Sustainable Development Goals (SDGs) and 5P

Partnership

Peace

Prosperity

People

Planet

Green Finance

Generic term implying use or diversion of financial resources to deploy and support projects with long term positive impact on the environment.
Sustainable Finance

Finances deployed in support of projects that ensure just, sustainable and inclusive growth or attainment of one or more sustainable development goals.

Carbon Finance

Financial instruments based on economic value of carbon emissions which an organization cannot avoid but which it offsets by funding other compensatory projects that contribute to carbon emissions reduction

Climate Finance

Finances deployed in support of low carbon and climate resilient projects that help in climate change mitigation and adaptation efforts, particularly in the energy and infrastructure sectors.

ESG Investing

Investments considering the broad range of environmental (e.g. climate change, pollution biodiversity loss), social (e.g. working conditions, human rights, salary or compensation structures) and governance (e.g. board composition, diversity and inclusion, taxes) characteristics of the projects or companies being invested in; ethical and business sustainability considerations are integral part of financing.

Impact Investing

Investing in projects that solve a social or environmental problem; the focus is on the positive impact rather than the means used to produce that impact.

Dynamic Trends of Green Finance and Energy Policy

ESG:

Environmental

Social

Governance
CSR: Corporate Social Responsibility
ESG to 17 SDGs

ESG to 17 SDGs

Generative AI for ESG Applications
## AI and Sustainability Development Goals (SDGs)

<table>
<thead>
<tr>
<th>SDGs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>No poverty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero hunger</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good health and well-being</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender equality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean water and sanitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordable and clean energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decent work and economic growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry, innovation and infrastructure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduces inequalities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustainable cities and communities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsible consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climate action</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life below water</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life on land</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace, justice and strong institutions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partnerships for the goals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Economic
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17

### Ecological
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17

### Social
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17

### Positive impact of AI*
- 100%
- 76%
- 69%
- 10%
- 56%
- 100%
- 100%
- 92%
- 100%
- 90%
- 100%
- 82%
- 80%
- 90%
- 100%
- 58%
- 26%

Note: ● adopted from Vinuesa et al. (2020), ○ added based on our analysis.

*The assessment of AI’s possible positive impact is based on a consensus-based expert elicitation process (Vinuesa et al., 2020).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Code characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary objective&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Develop new (AI) methods (11/95)</td>
</tr>
<tr>
<td></td>
<td>Compare (AI) methods (39/95)</td>
</tr>
<tr>
<td></td>
<td>Apply (AI) methods (53/95)</td>
</tr>
<tr>
<td></td>
<td>Develop new system (20/95)</td>
</tr>
<tr>
<td></td>
<td>Other objective (4/95)</td>
</tr>
<tr>
<td>Sustainability dimension</td>
<td>Economic (23/95)</td>
</tr>
<tr>
<td></td>
<td>Ecological (17/95)</td>
</tr>
<tr>
<td></td>
<td>Social (72/95)</td>
</tr>
<tr>
<td>Sustainable Development Goals (SDGs)</td>
<td>SDG 1 (0/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 2 (2/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 3 (85/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 4 (6/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 5 (0/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 6 (0/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 7 (9/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 8 (7/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 9 (8/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 10 (1/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 11 (9/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 12 (8/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 13 (2/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 14 (0/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 15 (2/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 16 (11/95)</td>
</tr>
<tr>
<td></td>
<td>SDG 17 (0/95)</td>
</tr>
<tr>
<td>Data source</td>
<td>Reviews (12/95)</td>
</tr>
<tr>
<td></td>
<td>Social media/Online forums (31/95)</td>
</tr>
<tr>
<td></td>
<td>Health records (21/95)</td>
</tr>
<tr>
<td></td>
<td>Environment/Weather (10/95)</td>
</tr>
<tr>
<td></td>
<td>Energy (5/95)</td>
</tr>
<tr>
<td>Data source plurality</td>
<td>Single source (50/95)</td>
</tr>
<tr>
<td></td>
<td>Multiple sources (44/95)</td>
</tr>
<tr>
<td></td>
<td>N/A (1/95)</td>
</tr>
<tr>
<td>Data sensitivity</td>
<td>Publicly available data (64/95)</td>
</tr>
<tr>
<td></td>
<td>Internal data (16/95)</td>
</tr>
<tr>
<td></td>
<td>Other (11/95)</td>
</tr>
<tr>
<td></td>
<td>N/A (9/95)</td>
</tr>
<tr>
<td>Manual labeling</td>
<td>Yes (32/95)</td>
</tr>
<tr>
<td></td>
<td>No (63/95)</td>
</tr>
<tr>
<td>Technology</td>
<td>ML (91/95)</td>
</tr>
<tr>
<td></td>
<td>NLP (42/95)</td>
</tr>
<tr>
<td></td>
<td>CV (12/95)</td>
</tr>
<tr>
<td></td>
<td>Other (21/95)</td>
</tr>
<tr>
<td>Type of learning for ML approach</td>
<td>Supervised learning (85/95)</td>
</tr>
<tr>
<td></td>
<td>Unsupervised learning (23/95)</td>
</tr>
<tr>
<td>Neural vs. non-neural</td>
<td>Non-neural (45/95)</td>
</tr>
<tr>
<td></td>
<td>Neural (50/95)</td>
</tr>
<tr>
<td></td>
<td>Deep learning (38/95)</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Technical evaluation (83/95)</td>
</tr>
<tr>
<td></td>
<td>Domain evaluation (25/95)</td>
</tr>
<tr>
<td>Paradigm</td>
<td>DSR/ADR (30/95)</td>
</tr>
<tr>
<td></td>
<td>Non-DSR/ADR (64/95)</td>
</tr>
<tr>
<td></td>
<td>0-9</td>
</tr>
<tr>
<td></td>
<td>10-29</td>
</tr>
<tr>
<td></td>
<td>30-54</td>
</tr>
<tr>
<td></td>
<td>55-69</td>
</tr>
<tr>
<td></td>
<td>70-95</td>
</tr>
</tbody>
</table>

Notes: Code dimensions are not mutually exclusive; one article can be classified into one or more code characteristics; <sup>1</sup> Compare does include ‘apply’.  

Sustainable Productivity: Finance ESG

FINANCE
- Financial result
- Financial expense

ENVIRONMENT
- Positive environmental contribution
- Environmental impact

SOCIAL
- Positive social contribution
- Social burden

GOVERNANCE
- Positive governance contribution
- Governance expense

Sustainable Resilient Manufacturing

ESG

**FINANCE**
- Sales revenue
- Equity ratio
- Value creation per employee

**GOVERNANCE**
- Innovation capability
- Renumeration structures
- Working conditions in the value chain

**ENVIRONMENTAL**
- Emissions impact
- Reuse quota
- Compensation measures

**SOCIAL**
- Gender equality
- Long-term view
- Integration & collaboration

ESG Indexes

• MSCI ESG Index
• Dow Jones Sustainability Indices (DJSI)
• FTSE ESG Index
MSCI ESG Rating Framework

**DATA**
1,000+ data points on ESG policies, programs, and performance; Data on 100,000 individual directors; up to 20 years of shareholder meeting results

**EXPOSURE METRICS**
How exposed is the company to each material issue? Based on over 80 business and geographic segment metrics

**MANAGEMENT METRICS**
How is the company managing each material issue? 150 policy/program metrics, 20 performance metrics; 100+ Governance Key Metrics

**SOURCE**
100+ specialized datasets (government, NGO, models)
Company disclosure (10-K, sustainability report, proxy report)
3,400+ media sources monitored daily (global and local news sources, governments, NGOs)

**KEY ISSUE SCORES & WEIGHTS**
35 Key Issues selected annually for each industry and weighted based on MSCI's materiality mapping framework.

**ESG RATING (AAA-CCC)**
Issue scores and weights combine to overall ESG rating relative to industry peers.
- **Individual E, S, G scores also available**

**INSIGHT**
Specialized ESG research team provides additional insight through:
- Company reports
- Industry reports
- Thematic reports
- Analyst calls & webinars

**MONITORING & QUALITY REVIEW**
Systematic, ongoing daily monitoring of controversies and governance events
In-depth quality review processes at all stages of rating, including formal committee review

**DATA OUTPUTS**
Access to selected underlying data
- Ratings, scores, and weights on 680,000 securities
- 17 years of history

### MSCI ESG Key Issue Hierarchy

<table>
<thead>
<tr>
<th>3 Pillars</th>
<th>10 Themes</th>
<th>35 ESG Key Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>Climate Change</td>
<td>Carbon Emissions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Product Carbon Footprint</td>
</tr>
<tr>
<td></td>
<td>Natural Capital</td>
<td>Water Stress</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biodiversity &amp; Land Use</td>
</tr>
<tr>
<td></td>
<td>Pollution &amp; Waste</td>
<td>Toxic Emissions &amp; Waste</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Packaging Material &amp; Waste</td>
</tr>
<tr>
<td></td>
<td>Environmental Opportunities</td>
<td>Opportunities in Clean Tech</td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>Human Capital</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Labor Management</td>
</tr>
<tr>
<td></td>
<td>Product Liability</td>
<td>Product Safety &amp; Quality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chemical Safety</td>
</tr>
<tr>
<td></td>
<td>Stakeholder Opposition</td>
<td>Consumer Financial Protection</td>
</tr>
<tr>
<td></td>
<td>Social Opportunities</td>
<td>Controversial Sourcing</td>
</tr>
<tr>
<td></td>
<td>Governance</td>
<td>Community Relations</td>
</tr>
<tr>
<td>Governance</td>
<td>Corporate Governance</td>
<td>Access to Communications</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Access to Finance</td>
</tr>
<tr>
<td></td>
<td>Corporate Behavior</td>
<td>Access to Health Care</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Opportunities in Nutrition &amp; Health</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ownership &amp; Control</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Board</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Business Ethics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tax Transparency</td>
</tr>
</tbody>
</table>

MSCI Governance Model Structure

Sustainalytics’ ESG Risk Ratings measure a company’s exposure to industry-specific material ESG risks and how well a company is managing those risks.

<table>
<thead>
<tr>
<th>Negligible</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 10</td>
<td>10 - 20</td>
<td>20 - 30</td>
<td>30 - 40</td>
<td>40+</td>
</tr>
</tbody>
</table>

Source: https://www.sustainalytics.com/esg-data
Truvalue Labs applies AI to analyze over 100,000 sources and uncover ESG risks and opportunities hidden in unstructured text.

The ESG Ranks data service produces an overall company rank based on industry percentile leveraging the 26 ESG categories defined by the Sustainability Accounting Standards Board (SASB).

The data feed covers 20,000+ companies with more than 13 years of history.

Source: https://developer.truvaluelabs.com/data/esg-ranks
Analyst-driven vs. AI-driven ESG

Analyst-driven ESG research
Derives ratings in a structured data model

1. Collect data
2. Process data
3. Analyze data
4. Generate score/rating

Data Miners & Technology
Analysts

Sustainalytics

Analyst role at the end of the process allows subjectivity to color results

AI-driven ESG research
Derives signals from unstructured data

1. Develop algorithms
2. Extract signals
3. Analyze data
4. Generate score/rating

Analysts & Engineers
Technology

Truvalue Labs

Analyst expertise at the beginning of the process produces consistent results

Analyst based ESG Research

AI based ESG Research

Analyst Based ESG Research

- Collect Data: Raw data sources and tools
- Process Data: Analysts
- Analyze Data: Analysts
- Generate score rating

Applying AI to ESG Research

- Develop algorithms: Analysts and Engineers
- Extract signals: Technology
- Analyze data and uncover material impact: Analysts
- Review and determine investment decision

It would take an analyst over 5 years to do what our AI can in 1 week
Combining analysts with AI creates gives you the full picture

Source: https://www.esganalytics.io/insights/how-data-is-accelerating-esg
ESG Analytics: NLP Taxonomy

UN SDG
SASB
CFAI
GRI

ESG Analytics Master View

Environmental
- Climate Change (5 subtopics)
- Sustainability (8 subtopics)
- Biodiversity & Water (3 subtopics)

Social
- Health & Safety (2 subtopics)
- Employee Standards (6 subtopics)
- Community Responsibility (7 subtopics)
- Human Rights (5 subtopics)

Governance
- Leadership & Management (4 subtopics)
- Business Innovation (3 subtopics)
- Outside Activities (4 subtopics)
- Business Ethics (4 subtopics)

Source: https://www.esganalytics.io/insights/how-data-is-accelerating-esg
**Top ESG Reporting Software**

Environmental, Social and Governance (ESG) Reporting software or Sustainability software helps organizations manage their operational data, evaluate their impact on the environment and provide reporting to perform audits.

ESG Reporting Software: Emitwise

• Emitwise is the carbon management platform for companies with complex manufacturing supply chains to confidently understand, track and reduce their complete carbon footprint.

• Combining 100 years of carbon accounting experience and machine learning technology, we accelerate climate action by increasing the accuracy of scope 3 emissions.

• The platform empowers manufacturers and their supply chains to make carbon-led business decisions that lower risk, increase profitability and deliver ambitious climate action.

ESG Reporting Software: Workiva ESG

• Workiva is a cloud native platform that simplifies the complexities of reporting and compliance.

• Workiva ESG is the end-to-end platform that allows you to integrate financial data, nonfinancial data, and XBRL.

• Workiva, the platform that streamlines your entire ESG process.

• Automate data collection, utilize frameworks, and directly connect to all your ESG reports. In meaningful glossy reports, accurate survey responses, and regulatory filings with integrated XBRL tagging.

8.4  8.7  +92  89%

COMPOSITE SCORE  CX SCORE  EMOTIONAL FOOTPRINT  LIKELINESS TO RECOMMEND

The Quant Finance PyData Stack

Quantopian

PyThalesians

Zipline

DX Analytics

PyAlgoTrade

QuantLib

PyTables

NetworkX

scikits-image

StatsModels

Statistics in Python

matplotlib

pandas

$y_t = \beta x_t + \mu_t + \epsilon_t$

SciPy

NumPy

Python

IPython

Jupyter
Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.
Denis Rothman (2021),
Transformers for Natural Language Processing:
Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more,
Packt Publishing.

Source: https://www.amazon.com/Transformers-Natural-Language-Processing-architectures/dp/1800565798
Savaş Yıldırım and Meysam Asgari-Chenaghlu (2021),

**Mastering Transformers:**
Build state-of-the-art models from scratch with advanced natural language processing techniques,

Packt Publishing.

Source: [https://www.amazon.com/Mastering-Transformers-state-art-processing/dp/1801077657/](https://www.amazon.com/Mastering-Transformers-state-art-processing/dp/1801077657/)

Source: https://www.amazon.com/Practical-Natural-Language-Processing-Pragmatic/dp/1492054054
Text Analytics and Text Mining

Text Mining “Knowledge Discovery in Textual Data”

- Document Matching
- Link Analysis
- Search Engines
- Information Retrieval
- POS Tagging
- Lemmatization
- Word Disambiguation

Text Analytics

Web Mining
- Web Content Mining
- Web Structure Mining
- Web Usage Mining

Data Mining
- Classification
- Clustering
- Association

Natural Language Processing

Management Science

Machine Learning

Computer Science

Artificial Intelligence

Statistics

Other Disciplines

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Generative AI

Text, Image, Video, Audio Applications
Generative AI (Gen AI)

AI Generated Content (AIGC)

Image Generation

**Instruction 1:**
An astronaut riding a horse in a photorealistic style.

**Instruction 2:**
Teddy bears working on new AI research on the moon in the 1980s.

Figure 1

Figure 2

Generative AI (Gen AI)
AI Generated Content (AIGC)

Unimodal

Please write a story about a cat.

Instruction $I_1$ → Prompt → Pre-train → Data → Generative AI Models → Decode → Result $R_1$

Multimodal

Describe this picture.

Instruction $I_2$ → Prompt → Pre-train → Data → Generative AI Models → Result $R_2$

Draw a picture of a cat.

Instruction $I_3$ → Prompt → Pre-train → Data → Generative AI Models → Result $R_3$

Write a song about a cat.

Instruction $I_4$ → Prompt → Pre-train → Data → Generative AI Models → Result $R_4$

Once upon a time, there was a cat named Jessy....

This is a cat.

The history of Generative AI in CV, NLP and VL

Unimodal - CV & NLP

- CV: GAN, VAE, Flow
- NLP: N-Gram, LSTM/GRU, Transformer
- VL: Show-Tell, StyleNet, StackGAN

- CV: BiGAN, RevNet
- NLP: ELMO, BERT, GPT-2
- VL: CAVP, DMGAN, VQ-VAE

- CV: StyleGAN, BigBiGAN
- NLP: GPT-3, OPT, BART, T5
- VL: VisualBERT, ViLBERT, UNITER

- CV: CLIP, ALBEF, BLIP, VQ-GAN
- NLP: DALL-E, BLIP2, DALL-E 2

Multimodal – Vision Language

Generative AI
Foundation Models

Categories of Vision Generative Models

(1) Generative adversarial networks

(2) Variational autoencoders

(3) Normalizing flows

(4) Diffusion models

The General Structure of Generative Vision Language

Two Types of Vision Language Encoders: Concatenated Encoders and Cross-aligned Encoders

(a) Concatenated Encoder

(b) Cross-aligned Encoder

Two Types of to-language Decoder Models: Jointly-trained Models and Frozen Models

(a) Jointly-trained Models

(b) Frozen Models

ChatGPT
Large Language Models (LLMs)
Foundation Models
Large Language Models (LLM) (GPT-3, ChatGPT, PaLM, BLOOM, OPT-175B, LLaMA)
The Transformers Timeline

Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.
Transformer Models

Transformer

Encoder

Decoder

DistilBERT

BERT

RoBERTa

XLM-R

XLM

ALBERT

ELECTRA

DeBERTa

T5

BART

M2M-100

BigBird

mT0

GPT

GPT-2

CTRL

GPT-3

GPT-Neo

GPT-J

BLOOM

BLOOMZ

ChatGPT

GPT-4

Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.
ChatGPT is not all you need. Attention is all you need.

Large Language Models (LLMs) (larger than 10B)

Ratios of various data sources in the pre-training data for existing LLMs

## Typical Data Preprocessing Pipeline for Pre-training Large Language Models (LLMs)

### Raw Corpus

- Alice is writing a paper about LLMs.
- Alice is writing a paper about LLMs.

### Quality Filtering

- Language Filtering
- Metric Filtering
- Statistic Filtering
- Keyword Filtering

### De-duplication

- Sentence-level
- Document-level
- Set-level

### Privacy Reduction

- Detect Personality Identifiable Information (PII)
- Remove PII

### Tokenization

- Reuse Existing Tokenizer
- SentencePiece
- Byte-level BPE

### Ready to pre-train!

32, 145, 66, 79, 12, 56, ...

---

The AI community building the future.

Build, train and deploy state of the art models powered by the reference open source in machine learning.

https://huggingface.co/
State-of-the-art Machine Learning for Jax, Pytorch and TensorFlow

 Transformers (formerly known as pytorch-transformers and pytorch-pretrained-bert) provides thousands of pretrained models to perform tasks on different modalities such as text, vision, and audio.

These models can applied on:

- Text, for tasks like text classification, information extraction, question answering, summarization, translation, text generation, in over 100 languages.
- Images, for tasks like image classification, object detection, and segmentation.
- Audio, for tasks like speech recognition and audio classification.

Transformer models can also perform tasks on several modalities combined, such as table question answering, optical character recognition, information extraction from scanned documents, video classification, and visual question answering.

https://huggingface.co/docs/transformers/index
## Hugging Face Tasks

### Natural Language Processing

<table>
<thead>
<tr>
<th>Task</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text Classification</td>
<td>3345</td>
</tr>
<tr>
<td>Token Classification</td>
<td>1492</td>
</tr>
<tr>
<td>Question Answering</td>
<td>1140</td>
</tr>
<tr>
<td>Translation</td>
<td>1467</td>
</tr>
<tr>
<td>Summarization</td>
<td>323</td>
</tr>
<tr>
<td>Text Generation</td>
<td>3959</td>
</tr>
<tr>
<td>Fill-Mask</td>
<td>2453</td>
</tr>
<tr>
<td>Sentence Similarity</td>
<td>352</td>
</tr>
</tbody>
</table>

[https://huggingface.co/tasks](https://huggingface.co/tasks)
Artificial Intelligence in Finance

About this Repository

This repository provides Python code and Jupyter Notebooks accompanying the Artificial Intelligence in Finance book published by O’Reilly.
Yves Hilpisch (2020), *Artificial Intelligence in Finance: A Python-Based Guide*, O’Reilly


<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>directory.oanda</td>
<td>Code updates for TF 2.3.</td>
</tr>
<tr>
<td>01_artificial_intelligence.ipynb</td>
<td>Code updates for TF 2.3.</td>
</tr>
<tr>
<td>02_superintelligence.ipynb</td>
<td>Code updates for TF 2.3.</td>
</tr>
<tr>
<td>03_normative_finance.ipynb</td>
<td>Code updates for TF 2.3.</td>
</tr>
<tr>
<td>04_data_driven_finance_a.ipynb</td>
<td>Initial commit.</td>
</tr>
<tr>
<td>04_data_driven_finance_b.ipynb</td>
<td>Initial commit.</td>
</tr>
<tr>
<td>05_machine_learning.ipynb</td>
<td>Code updates for TF 2.3.</td>
</tr>
<tr>
<td>06_ai_first_finance.ipynb</td>
<td>Code updates for TF 2.3.</td>
</tr>
<tr>
<td>07_dense_networks.ipynb</td>
<td>Code updates for TF 2.3.</td>
</tr>
<tr>
<td>08_recurrent_networks.ipynb</td>
<td>Code updates for TF 2.3.</td>
</tr>
<tr>
<td>09_reinforcement_learning_a.ipynb</td>
<td>Code updates.</td>
</tr>
<tr>
<td>09_reinforcement_learning_b.ipynb</td>
<td>Code updates for TF 2.3.</td>
</tr>
</tbody>
</table>

Source: [https://github.com/yhilpisch/aiif/tree/main/code](https://github.com/yhilpisch/aiif/tree/main/code)
Python in Google Colab (Python101)

https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT

---

```python
# Future Value
pv = 100
r = 0.1
n = 7
fv = pv * (1 + (r)) ** n
print(round(fv, 2))
```

194.87

```python
amount = 100
interest = 10 #10% = 0.01 * 10
years = 7
future_value = amount * ((1 + (0.01 * interest)) ** years)
print(round(future_value, 2))
```

194.87

```python
# Python Function def
def getFv(pv, r, n):
    fv = pv * (1 + (r)) ** n
    return fv
fv = getFv(100, 0.1, 7)
print(round(fv, 2))
```

194.87

```python
# Python if else
score = 80
if score >=60 :
    print("Pass")
else:
    print("Fail")
```

Pass

---

https://tinyurl.com/aintpupython101
Teaching

• Artificial Intelligence in Finance and Quantitative
  • Fall 2021, Fall 2022, Fall 2023
• Artificial Intelligence for Text Analytics
  • Spring 2022, Fall 2023
• Big Data Analytics
  • Fall 2020, Spring 2023
• Software Engineering
  • Fall 2020, Fall, 2021, Spring 2022, Spring 2023
• Artificial Intelligence
  • Spring 2021, Fall 2022
• Data Mining
  • Spring 2021
• Foundation of Business Cloud Computing
  • Spring 2021, Spring 2022, Spring 2023
• Python for Accounting Applications
  • Fall 2023

https://web.ntpu.edu.tw/~myday/teaching.htm
1. Applying AI technology to construct knowledge graphs of cryptocurrency anti-money laundering: a few-shot learning model
   • MOST, 110-2410-H-305-013-MY2, 2021/08/01~2023/07/31
2. Fintech Green Finance for Carbon Market Index, Corporate Finance, and Environmental Policies. Carbon Emission Sentiment Index with AI Text Analytics
   • NTPU, 112-NTPU_ORDA-F-003, 2023/01/01~2024/12/31
   • NSTC 112-2425-H-305-002-, 2023/05/01-2026/04/30
4. Establishment and Implement of Smart Assistive Technology for Dementia Care and Its Socio-Economic Impacts. Intelligent, individualized and precise care with smart AT and system integration
   • NSTC, NSTC, 112-2627-M-038-001-, 2023/08/01~2024/07/31
5. Use deep learning to identify commercially dental implant systems - observational study
   • USTP-NTPU-TMU, USTP-NTPU-TMU-112-01, 2023/01/01~2023/12/31
6. Metaverse AI Multimodal Cross-Language Task-Oriented Dialogue System
   • ATEC Group x NTPU, NTPU-112A413E01, 2023/05/01~2026/04/30
7. Metaverse Avatar Automatic Metadata Generation Module
   • FormosaVerse x NTPU, NTPU-111A413E01, 2022/12/01~2023/11/30
8. Pilot Study on Universal Data Processing for Code Generation Engine
   • III x NTPU, NTPU-112A513E01, 2023/08/01~2023/12/22

https://web.ntpu.edu.tw/~myday/cindex.htm#projects
Summary

• This course introduces the fundamental concepts and hands-on practices of Python for Accounting Applications.

• Topics include
  • Introduction to Python for Accounting Applications,
  • Python Programming and Data Science,
  • Foundations of Python Programming,
  • Data Structures, Control Logic and Loops,
  • Functions and Modules,
  • Files and Exception Handling,
  • Data Analytics and Visualization with Python,
  • Obtaining Data From the Web with Python,
  • Statistical Analysis with Python, Machine Learning with Python,
  • Text Analytics with Python with LLMs,
  • Applications of Accounting Data Analytics with Python, and
  • Applications of ESG Data Analytics with Python.
Contact Information

Min-Yuh Day, Ph.D.
Associate Professor
Institute of Information Management, National Taipei University
Tel: 02-86741111 ext. 66873
Office: B8F12
Address: 151, University Rd., San Shia District, New Taipei City, 23741 Taiwan
Email: myday@gm.ntpu.edu.tw
Web: http://web.ntpu.edu.tw/~myday/