**1. Project Name**

Country Offices Programme Documents (CPD) Portfolio Sensemaking

**2. Overview**

Portfolio sensemaking helps UNDP Country Offices (COs) to establish the strategic relevance of their portfolio of projects, enhance portfolio coherence, and over time develop and incorporate sensemaking as an institutional capacity at the core. That is the reason UNDP Regional Innovation Centre (RIC) in Bangkok had an open call for consultation for “Portfolio analysis for strategic insights”. This workstream aims to identify patterns, connections, similarities and differences between UNDP's projects on regional and country-level using Natural Language Processing (NLP) tools applied on projects documents. With a structured process of deep-reflection and socialization, portfolio sensemaking is expected to help COs to establish the strategic relevance of their portfolio of projects. The overall guideline for the workstream would be consistent with the UNDP RIC Bangkok call for the portfolio sensemaking.

**3. Background**

With a large number of ongoing and past projects, UNDP is one of the leading development organizations with wide spectrum of project portfolio. However, tracing why a project was implemented, if it intends to reflect the reality, what can be learnt from other past and present projects, where the intersections are and how resources can be bundled to increase efficiency, is an ongoing research question that frequently raised within the organization. In the past, qualitative approaches have been applied to tackle these issues, however, with the advancements in Machine Learning and specifically Natural Language Processing a quantitative dimension shall be added to gather new insights and benefit from augmentation. Complementing to the qualitative process of portfolio sensemaking, a large volume of unstructured data laying in the project documents and administrative reports could potentially be extracted and utilized for exploring the hidden connections and patterns. Most of the information are openly accessible on the official UNDP website, therefore barrier for experimentation is low.

**4. Key Objectives & Research Questions**

* Explore and identify the hidden connections and patterns among UNDP Country Office’s projects with the unstructured data in the organization.
* Adopt a data-driven approach to provide useful intelligence for portfolio sensemaking that aims at enhancing portfolio coherence and accelerating institutional learning. Create deliverable knowledge products for the process.
* Enhance the UNDP Country Office’s knowledge and capacity in gaining a basic understanding of applying Artificial Intelligence for development work, including how it works, what it can and cannot do, potential bias and limitation and implication of ethical use of AI.

Initial Iteration:

* Process and structure raw data in a machine-readable format.
* Apply unsupervised learning algorithms (e.g. topic modelling, clustering) to obtain thematic patterns in UNDP's projects.
* Visualize connections and gaps between projects to communicate the findings with non-expert readers.

Subsequent Iteration:

* Expand analysis from regional to country level documents and find patterns, gaps, connections between projects using unsupervised learning algorithms. Apply the same workflow from the previous iteration to optimize the result.
* Apply supervised learning algorithms, such as text classification to link projects to pre-defined categories. This requires annotation of project portfolio documents.
* Visualize the findings and create a communication material under SDG AI Lab label.

Throughout the work, the research questions that need to be answered are the following:

RQ1: What components are required to deploy the results and interactive charts from the first iteration into a website-based application?

RQ2: What is the best method of topic modelling technique to extract the hidden topics from document inputs and find patterns, gaps, connections between projects?

RQ3: What are underlying common topics we can find in Country Project Documents, according to each country?

RQ4: What are common topics of the Program rationales extracted from the CPD?

RQ5: What countries form close clusters regarding their project similarities? And what do those clusters have in common?

RQ6: How can this data pipeline benefit the Portfolio Sensemaking and Acceleration Protocol?

Please note that these suggestions are first and foremost guidelines. Consider the objectives as the focus, but may choose other methodologies and models that deems more suitable or accurate for the tasks.

**5. Workflows Ideas and Suggestions**

* Extract text and relevant data from the project document and project progress reports (in the format of scanned PDF) against the structure of the data frame.
* Conduct text mining, network analysis/topological analysis and other relevant data analysis with the support of machine learning (the models should be built for the UNDP Philippines’s context) to identify and visualize connections, patterns and clusters among projects based on key dimensions such as levers of change, thematic areas, partnerships, capacities and others.
* Develop quick prototypes to test out different types of possible data analysis and test them. The final data analysis should be captured in an interactive visual dashboard.
* Develop a brief research report to capture the key findings and learnings with visualization and an associated blog post on the Medium platform.

**6. Dataset and File Description**

Asia and the Pacific Regional and Country Programme Documents are available via this [link](https://www.asia-pacific.undp.org/content/rbap/en/home/about-us/legal-framework.html). However this is an example dataset as there are larger sets of data available on the official website.

|  |  |  |
| --- | --- | --- |
| **Country** | **Year** | **Programme Document** |
| Asia and the Pacific | [2018─2021](http://undocs.org/DP/RPD/RAP/3) | Regional Programme Document |
| Pacific Islands Countries and Territories:       The Cook Islands, Federated States of Micronesia, Fiji, Kiribati, Marshall Islands, Nauru, Niue,       Palau, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu and Vanuatu | [2018─2022](http://undocs.org/DP/DSP/PIC/2) | Sub-regional Programme Document |
| Afghanistan | [2015─2019](http://undocs.org/DP/DCP/AFG/3) | Country Programme Document |
| Bangladesh | [2017─2020](http://undocs.org/DP/DCP/BGD/3) | Country Programme Document |
| Bhutan | [2014─2018](https://www.asia-pacific.undp.org/content/dam/rbap/docs/programme-documents/BT-CP-2014-2018.pdf) [2019─2023](http://undocs.org/DP/DCP/BTN/2) | Country Programme Document |
| Cambodia | [2019─2023](https://www.kh.undp.org/content/dam/cambodia/docs/LatetstCPD_Sept/Cambodia%20CPD%202019-23_approved.pdf) | Country Programme Document |
| China | [2016─2020](http://undocs.org/DP/DCP/CHN/3) | Country Programme Document |
| DPR Korea | [2011─2015](https://www.asia-pacific.undp.org/content/dam/rbap/docs/programme-documents/KP-CP-2011-2015.pdf) | Country Programme Document |
| India | [2018─2022](http://undocs.org/DP/DCP/IND/3) | Country Programme Document |
| Indonesia | [2016-2020](http://undocs.org/DP/DCP/IDN/3) [2021-2025](https://www.asia-pacific.undp.org/content/dam/rbap/docs/programme-documents/CPD-Ind_2021-2025.DOC) | Country Programme Document |
| Iran | [2017─2021](http://undocs.org/DP/DCP/IRN/3) | Country Programme Document |
| Lao PDR | [2017─2021](http://undocs.org/DP/DCP/LAO/3) | Country Programme Document |
| Malaysia | [2016─2020](http://undocs.org/DP/DCP/MYS/3) | Country Programme Document |
| Maldives | [2016─2020](http://undocs.org/DP/DCP/MDV/3) | Country Programme Document |
| Mongolia | [2017─2021](http://undocs.org/DP/DCP/MNG/3) | Country Programme Document |
| Myanmar | [2018─2022](http://undocs.org/DP/DCP/MMR/2) | Country Programme Document |
| Nepal | [2018─2022](http://undocs.org/DP/DCP/NPL/3) | Country Programme Document |
| Pakistan | [2018─2022](http://undocs.org/DP/DCP/PAK/2) | Country Programme Document |
| Papua New Guinea | [2018─2022](http://undocs.org/DP/DCP/PNG/2) | Country Programme Document |
| Philippines | [2012─2016](https://www.asia-pacific.undp.org/content/dam/rbap/docs/programme-documents/PH-CP-2012-2016.pdf) [2019─2023](http://undocs.org/DP/DCP/PHL/3) | Country Programme Document |
| Sri Lanka | [2018─2022](http://undocs.org/DP/DCP/LKA/3) | Country Programme Document |
| Thailand | [2017─2021](http://undocs.org/DP/DCP/THA/3) | Country Programme Document |
| Timor-Leste | [2015-2019](http://undocs.org/DP/DCP/TLS/2) [2021-2025](https://www.asia-pacific.undp.org/content/dam/rbap/docs/programme-documents/CPD_TL-2021-2025.DOC) | Country Programme Document |
| Viet Nam | [2017─2021](http://undocs.org/DP/DCP/VNM/2) | Country Programme Document |

**7. Deliverables**

Initial Iteration:

* Data segmentation and processing for further analysis
* Text mining, network analysis/topological analysis and other relevant data analysis to identify and visualize connections, patterns or break of them, clusters among projects based on key dimensions

Subsequent Iteration:

* The same list of deliverable applied with the data provided by a partner agency.
* A research brief under SDG AI Lab label