Background

Client Organization

The State of Michigan - Department of Technology, Management and Budget (DTMB) plays a pivotal role in supporting the business operations of various state agencies in Michigan through a range of services, including building management, information technology, centralized contracting, budget management, space planning, construction oversight, fleet operations, and retirement system administration. Central Procurement Services (CPS) is a division within this department responsible for overseeing state purchases for all agencies in Michigan.

Central Procurement Services (CPS) manages the acquisition of goods and services for state agencies, ensuring efficiency and compliance with state regulations. CPS's responsibilities include developing procurement policies, streamlining purchasing processes, and fostering a competitive business environment. In essence, CPS aims to maintain transparency, integrity, and provide the best value for taxpayers' money.

Project Context

In Michigan, the Central Procurement Office faces an issue with time-consuming and error-prone manual data extraction from state contracts. Currently, there is no streamlined process to extract crucial data, such as contract numbers, NIGP codes, and detailed descriptions of commodities or services from a large volume of contracts.

This lack of an efficient system affects various stakeholders, including state agencies, who face challenges in quickly identifying relevant contracts. This problem affects state procurement officials, taxpayers, and state agencies by draining resources, introducing errors, and causing delays. The lack of a comprehensive database impedes the procurement process, often leading to delays and increased costs for the state.

This project aims to automate this process, resulting in time and cost savings, improved accuracy, quicker decision-making, and increased transparency. This project seeks to enhance efficiency and effectiveness in state procurement, benefiting both public finances and service delivery. If this problem is not addressed, the State of Michigan may continue to experience inefficiencies in the procurement process, resulting in wasted time and resources. This project aligns with CPS's mission of improving the efficiency of procurement processes in Michigan.

Relevant Stakeholders

- Procurement Officials: These officials are responsible for managing state contracts and procurement processes. The project aims to make their tasks more efficient and effective.
- Taxpayers: The project's success can lead to cost savings, ensuring that their tax dollars are used more efficiently.
- State Agencies: Various state agencies are primary users of the procurement system. The project will directly impact state agencies' ability to efficiently access and utilize state contracts.

Problem Statement

The primary context of this project is to improve the search functionality for commodities and services within the state contracts dashboard. The project's ultimate goal is to save time and money for the State of Michigan by streamlining the procurement process.

Project Objectives

Primary Objective: The primary objective of this project is to enhance operational efficiency by improving the search feature within the state contracts database. This improvement aims to make the search process faster and more efficient, benefiting stakeholders by saving them time and effort in locating contracts.

Secondary Objective: The secondary objective is to achieve significant cost and time savings, allowing stakeholders to focus on their core responsibilities rather than spending valuable time on the manual and time-consuming task of searching for contracts.

Assumptions:

Data Availability: The project team assumes access to the dataset containing the state contracts, which is expected to be in an electronic format and accessible throughout the project's duration.

Computational Resources: The team assumes access to essential computational resources, including high-performance computers and GPUs.

Project Team Expertise: It is assumed that, while the team may lack prior experience in machine learning and data scraping from PDFs, they have the capability to learn and effectively apply these skills.

Remote Project Execution: The team assumes the ability to collaborate effectively and complete the project successfully without the need for a physical presence at CPS's on-site location.

Constraints:

Lack of Prior Machine Learning Experience: The project team has not previously worked on machine learning projects in Python. This constraint may require additional time for learning and may result in a steeper learning curve.

Inexperience in Data Scraping from PDFs: The team does not have prior experience in extracting data from PDF documents. Overcoming this constraint will involve gaining expertise in PDF scraping methods.

Limited Project Timeline: The project is expected to be completed within a specific timeframe. This timeline constraint may limit the extent of complexity and sophistication that can be achieved within the project.

Dependence on Existing Data: The project depends on the quality and availability of the existing dataset. Any limitations in the dataset may impact the quality of the machine-learning model.

Anticipated Benefits and Impacts

The successful completion of this project is expected to yield several benefits and impacts:

- Improved efficiency in the procurement process for state agencies.
- Reduction in the time required to search for and identify relevant commodities and services within contracts.
- Cost savings for the State of Michigan.
- Enhanced data management capabilities for CPS.

- Potential for scalability and application to other areas within the Department of Technology, Management, and Budget.

This project is an essential step towards the State of Michigan's goal of improving transparency, efficiency, and value in its procurement processes. By addressing the problem of inefficient data extraction, this project aims to bring about positive changes that benefit CPS and the relevant stakeholders.

Project Resources and Needs

Research Goal

The research project aims to address the problem of time-consuming and error-prone manual data extraction from state contracts faced by the Central Procurement Office (CPS) in Michigan. The primary goal of this research project is to develop and implement a robust data extraction system for state contracts, using unsupervised machine learning techniques. This system aims to streamline the search for commodities and services within the contracts, ultimately saving time and money for the State of Michigan. To address this challenge, this research project will focus on automating two crucial components of the contract data extraction process. These components are integral to the automation pipeline and will significantly contribute to the goal of improving the efficiency of procurement processes in Michigan.

Automating PDF Download:

As the first task, the project involves developing a PDF download function that will automate the process of acquiring contract PDFs. This function will navigate through the CPS dashboard, follow links to the relevant contract PDFs, and download them to a designated folder. The automation of this process will significantly reduce the manual effort required to gather contract documents, resulting in substantial time savings and a decrease in potential errors.

Automating PDF to Text Conversion:

Once the PDFs are downloaded, the second function will convert these PDFs into a text format that can be analyzed. This function is essential for data extraction as it transforms PDF content into a machine-readable format. The resulting text can be processed, analyzed, and categorized efficiently, contributing to quicker decision-making and enhanced filtration.

The combination of these functions and the subsequent phases will help the team achieve the overarching goal of streamlining the procurement process, benefiting various stakeholders and saving valuable time and resources for the State of Michigan. This research project is crucial for improving transparency, efficiency, and value within the state's procurement operations.

Necessary Resources

Stakeholders/Internal Data

In order to complete the project, the team will not require extensive interaction with users, staff, or stakeholders. The primary contact, Mohamed Gibril, has provided comprehensive information regarding the project's scope and data access. However, given the team's relatively new collaboration with Mohamed, there may be instances where additional State of Michigan resources beyond his expertise may become necessary.

External Data

Regarding external data, this project entails the retrieval of external links for potential projects and clients. However, this external data is already integrated into our dataset. Therefore, we do not anticipate a significant need for additional external data, as it is unlikely to provide relevant information or enhance the project.

Skill Gap

There are several different skills that the team will need to learn before being able to complete the State of Michigan's final request. In order to be able to sift through data that is provided on a State of Michigan website, the team will need to become proficient in data scrapping. Specifically, the team will need to learn how to extract data from a PDF and be able to use this information to give better search results to individuals searching through State of Michigan contracts. Additionally, the team is going to need to become proficient in NLP and Machine Learning as these techniques will be used to extract labels and information from contracts within PDF files.

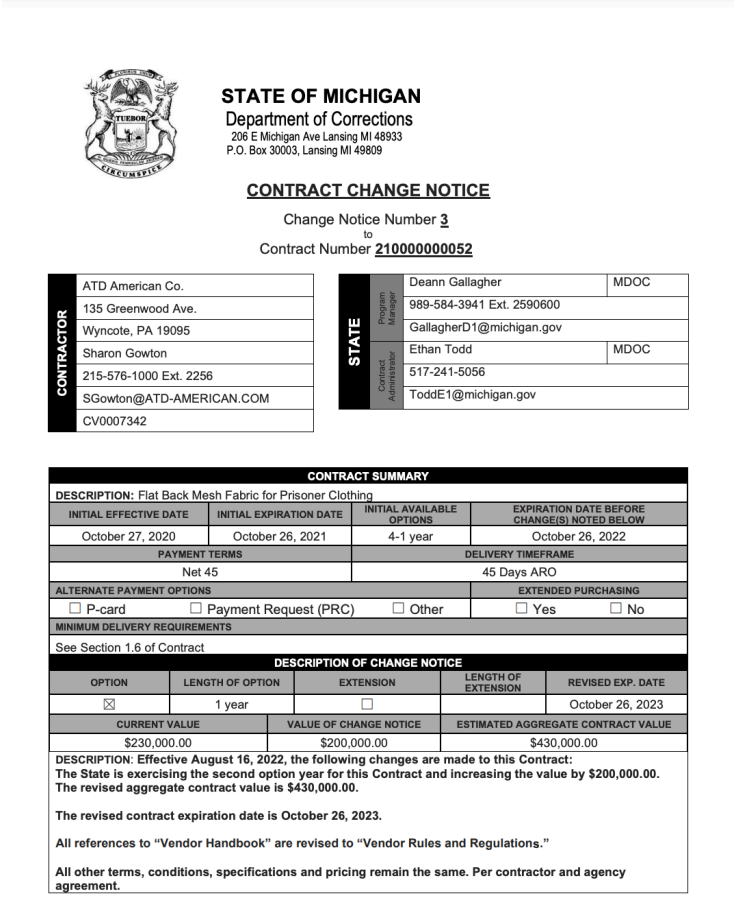
Identified data sets

The initial significant milestone for the project involves the development of a Python script to extract data from the existing Power BI dashboard pictured below. The Contract List serves as a directory for all contracts that are managed by State of Michigan Procurement. There are no specific restrictions or timelines associated with accessing this data. It's publicly available, and there are no requirements for passwords, request forms, or permissions to access it.

Contract Number	Description	Vendor Name	Link	Value	Start Date	Expiration Date
210000000052	Flat Back Mesh for Prisoner Clothing	A T D AMERICAN CO.		$430,000	10/27/2…	10/26/2023
210000000051	Linens, Blankets, Towels, and Fabri…	GRAND STRATEGY, LLC		$3,918,050	10/27/2…	10/26/2023
071B4300013	DRIVER'S LICENSE ID CARD	IDEMIA IDENTITY & SECURITY USA LLC		$67,823,022.79	3/18/2009	10/30/2023
190000001017	Captive Consulting Services	DICKINSON WRIGHT PLLC		$100,000	11/1/2017	10/31/2023
190000000044	Janitorial Services for the MDOT …	R & K SERVICES DBA SERVICEMASER HOME AND OFFICE		$83,434.9	11/1/2018	10/31/2023
220000000055	MDARD Geagley Laboratory Janit…	BOLING JANITORIAL SERVICE INC		$90,000	11/1/2021	10/31/2023
220000000056	MDARD Heffron Laboratory Janit…	BOLING JANITORIAL SERVICE INC		$62,000	11/1/2021	10/31/2023
190000000194	Snow and Ice Removal for MDOT …	OLSONS LAWNSCAPE, LLC		$72,750	11/1/2018	10/31/2023
200000000854	Electrician Support and Emergency…	J. RANCK ELECTRIC, INC.		$155,520	5/1/2020	10/31/2023
200000000078	Janitorial Services - Lansing CSC a…	HI-TEC BUILDING SERVICES INC		$76,285	11/1/2019	10/31/2023
200000000319	Ice for resale at various state parks	SUPERIOR ICE DISTRIBUTING, INC.		$93,342.5	2/10/2020	10/31/2023
200000000320	Ice for resale at various state parks	HOME CITY ICE CO.		$654,708.6	2/10/2020	10/31/2023
200000000321	Ice for resale at various state parks	ARCTIC GLACIER		$98,532	2/10/2020	10/31/2023
210000000106	New Ice Control Sand Contract	DOUBLE A INSPECTIONS, LLC		$37,232.75	11/1/2020	10/31/2023
220000001331	Survivor Moms Companion Training	GROWING FORWARD TOGETHER		$200,000	10/7/2022	10/31/2023
210000000064	Mecosta Osceola DHHS Snow Re…	FRED MYERS EXCAVATING & TRUCKING LLC		$58,636	11/1/2020	10/31/2023
071B3200144	MICROCUVETTES TEST KITS	RADIOMETER AMERICA INC		$3,730,125	11/11/2…	11/1/2023
240000000047	MDHHS Legal Affairs - Case Man…	COURTALERT.COM INC.		$0	10/11/2…	11/8/2023
210000001532	Signage Standardization Manual D…	KERESTES-MARTIN ASSOCIATES, INC		$63,351.5	11/15/2…	11/14/2023
180000001114	As-Needed Drainage Structure Rep…	GREAT LAKES CONTRACTING SOLUTIONS		$28,619,993.4	8/15/2018	11/14/2023
210000000109	Administrations, Scoring and Repo…	ACT INC		$14,799,000	11/15/2…	11/14/2023
190000000182	Water Well Testing for MDOT Bay…	QUADSIL INC		$111,910	11/19/2…	11/18/2023

Last Refresh: 10/24/2023

The dataset consists of seven columns and approximately 600 rows, each corresponding to an individual contract (ie. there are about 600 contracts). The column names in the dataset are “Contract Name,” “Description,” “Vendor Name,” “Link,” “Value,” “Start Date,” and “Expiration Date.” Of particular importance is the “Link” column, which contains hyperlinks to the aforementioned PDF government contracts. For example, when clicking on the hyperlink associated with Contract Number 210000000052, the following PDF contract is opened.



STATE OF MICHIGAN
Department of Corrections
206 E Michigan Ave Lansing MI 48933
P.O. Box 30003, Lansing MI 49809

CONTRACT CHANGE NOTICE

Change Notice Number **3**
to
Contract Number **210000000052**

CONTRACTOR	
	ATD American Co.
	135 Greenwood Ave.
	Wyncote, PA 19095
	Sharon Gowton
	215-576-1000 Ext. 2256
	SGowton@ATD-AMERICAN.COM
	CV0007342

STATE			
	Program Manager	Deann Gallagher	MDOC
		989-584-3941 Ext. 2590600	
		GallagherD1@michigan.gov	
	Contract Administrator	Ethan Todd	MDOC
		517-241-5056	
		ToddE1@michigan.gov	

CONTRACT SUMMARY

DESCRIPTION: Flat Back Mesh Fabric for Prisoner Clothing

INITIAL EFFECTIVE DATE	INITIAL EXPIRATION DATE	INITIAL AVAILABLE OPTIONS	EXPIRATION DATE BEFORE CHANGE(S) NOTED BELOW
October 27, 2020	October 26, 2021	4-1 year	October 26, 2022

PAYMENT TERMS	DELIVERY TIMEFRAME
Net 45	45 Days ARO

ALTERNATE PAYMENT OPTIONS			EXTENDED PURCHASING	
☐ P-card	☐ Payment Request (PRC)	☐ Other	☐ Yes	☐ No

MINIMUM DELIVERY REQUIREMENTS

See Section 1.6 of Contract

DESCRIPTION OF CHANGE NOTICE

OPTION	LENGTH OF OPTION	EXTENSION	LENGTH OF EXTENSION	REVISED EXP. DATE
☒	1 year	☐		October 26, 2023

CURRENT VALUE	VALUE OF CHANGE NOTICE	ESTIMATED AGGREGATE CONTRACT VALUE
$230,000.00	$200,000.00	$430,000.00

DESCRIPTION: Effective August 16, 2022, the following changes are made to this Contract:
The State is exercising the second option year for this Contract and increasing the value by $200,000.00.
The revised aggregate contract value is $430,000.00.

The revised contract expiration date is October 26, 2023.

All references to "Vendor Handbook" are revised to "Vendor Rules and Regulations."

All other terms, conditions, specifications and pricing remain the same. Per contractor and agency agreement.

The team's approach will involve web scraping this data into a structured pandas dataframe. Although the data is predominantly text, it lacks structured formatting, necessitating Natural Language Processing (NLP) manipulation to categorize the text. Besides the "link" column, the dataframe will incorporate the remaining columns such as "Contract Name," "Description," "Vendor Name," "Value," "Start Date," and "Expiration Date" to serve as supplementary identifiers for users to utilize when searching for specific contracts. This means that users can search for contracts not only based on the keywords extracted from the document but also by "Contract Name," "Description," "Vendor Name," and more.

External Data Resources

The project requires the text data within the linked contract PDFs in the dataset. The project team will need to write code to extract keywords from these descriptions and add them as a column in the dataframe.

Data Characteristics
There are no parameters consistent across the PDFs, but commonly there are contract summary and description sections that the team will scrape to extract keywords for the new dataframe column. The data within these PDFs is poorly structured and inconsistent. The PDFs are sometimes hundreds of pages with little standardization of sections or formatting. Despite the lack of structure, the data is complete as there is a multitude of information contained within these PDFs on each contract.

Data Scraping and Implementation
To access this data the team will develop a script to download the PDFs from the dashboard then use NLP and web scraping techniques to gather key words from the data. This data will be added to the current procurement data on the dashboard as an additional column. It will be merged with the current data on the contract number column. For this data to prove useful, the team would need to prove that the scraped data will support search and filter functionality within the dashboard, in order for stakeholders to find specific contracts quicker.

Timeline
There is no urgent timeline on the CPS side since the dashboard is currently functional, yet inefficient. The only time constraint is for the team to finish the project before the end of the academic year. The project team is responsible for completing this task of scraping and analyzing external data from the PDFs.

Skills to Learn

Skill: Data Scraping
Data scraping, often referred to as web scraping, is a technique used to extract information from websites. It automates the process of collecting data from the web, which otherwise would have to be done manually. For the project concerning the State of Michigan's request, data scraping is essential as the team needs to sift through data on their website, extract the necessary information, and PDF attachments to provide enhanced search results.

Proficiency in data scraping will streamline data collection, thus improving project efficiency. Learning data scraping can be accomplished through online courses, tutorials, or hands-on projects. Resources like Coursera, Udemy, or scraping tools like BeautifulSoup can be beneficial. The process to meet this need includes identifying the best learning resources, allocating time for team members to learn and practice, engaging in hands-on projects to reinforce learning, and if

necessary, consulting with the State of Michigan or seeking external expertise for advanced challenges.

We aim to achieve proficiency in data scraping by mid-December, considering the complexity of the required scraping tasks. This target provides a specific timeframe for completing the skill acquisition. The team has a clear allocation of responsibilities for skills and resources, with Jacob Hayes and Jason Kemp designated to concentrate on becoming proficient in data scraping, with Jacob Hays taking the lead in this area.

In case the team faces difficulties with web scraping, the backup plan involves two key strategies. Firstly, the team will explore user-friendly web scraping tools to simplify the process and reduce the learning curve. Secondly, the team will consider consulting with subject matter experts who have extensive experience in web scraping. These experts will provide guidance, assist in overcoming challenges, and potentially take on specific web scraping tasks.

Skill: Natural Language Processing (NLP) and Machine Learning (ML)
Natural Language Processing (NLP) and Machine Learning (ML) are pivotal branches of artificial intelligence. NLP focuses on the interaction between computers and human language, while ML provides systems the ability to learn from and improve with experience. NLP and ML are integral for extracting labels and information from contracts within PDF files, which is a crucial aspect of fulfilling the project's objectives. They will enable the automation and enhancement of search capabilities within the State of Michigan's contracts.

Acquiring skills in NLP and ML can be similar to data scraping, where team members can engage in online courses, tutorials, and hands-on projects. Platforms like Coursera, Udemy, or even specialized boot camps can be instrumental. The process to meet this need includes identifying valuable learning resources, allocating time for learning and practicing, working on small projects to apply the newly acquired skills, and seeking help from the State of Michigan or external expert advice for advanced challenges if necessary.

The team aims to achieve proficiency in NLP and ML by mid-December, considering the complexity of the required scraping tasks. This target provides a specific timeframe for completing the skill acquisition. The team has a clear allocation of responsibilities for skills and resources, with Abhinav Tadikonda and Micah Blackburn designated to concentrate on becoming proficient in in NLP and ML, with Abhinav Tadikonda taking the lead in this area.

If the team encounters challenges in mastering NLP and machine learning, the backup plan involves two primary strategies. Firstly, the team will explore the use of pre-built machine learning models and NLP libraries that can be readily applied to the project. These models can

simplify the development process and reduce the need for in-depth knowledge about the subject. Secondly, the team will consider consulting with subject matter experts who specialize in NLP and machine learning. These experts will provide guidance, help overcome specific challenges, and offer insights into best practices.

Wider Context

The initiative to automate data extraction for the State of Michigan - Department of Technology, Management, and Budget (DTMB) is part of a larger trend towards digital transformation in various sectors. Governments and organizations globally are harnessing innovative technologies to streamline operations, refine decision-making, and enhance service delivery. By implementing advanced data processing capabilities, these entities can significantly diminish manual workloads, reduce errors, and utilize data for strategic insights.

Comparative Analysis
The comparative analysis conducted by the team aims to identify parallels between the challenges faced by the Central Procurement Service (CPS) of the State of Michigan and solutions adopted in analogous scenarios. This examination provides a benchmark for assessing the effectiveness of various tools and methodologies in managing data extraction and handling. It involves evaluating the efficiency, precision, and cost benefits of diverse strategies, which can serve as a guide for the team's project.

Direct Comparators
Direct comparators include state procurement systems elsewhere that have embraced automated data extraction technologies. For instance, the California Department of General Services has adopted AI-based systems within its procurement operations, leading to a marked decrease in manual data processing. Insights from these implementations offer valuable perspectives on the practical application of such technologies within government settings.

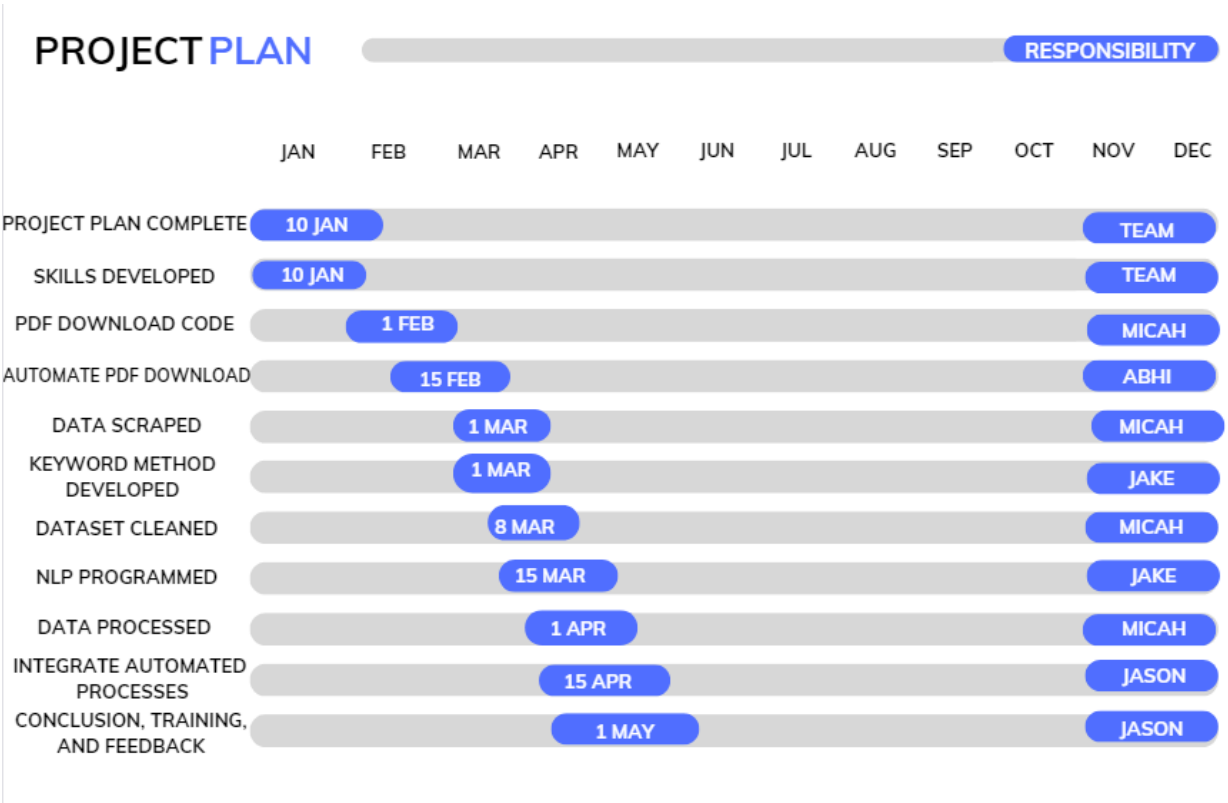
Indirect Comparators
Indirect comparators, such as private sector procurement systems like SAP Ariba, provide an alternate viewpoint on addressing similar issues. These systems employ machine learning algorithms to refine data extraction accuracy and operational efficacy. Though from a different sector, these solutions present innovative methodologies that can be tailored to the needs of the Central Procurement Service (CPS).

Analogous Research
Analogous research explores industries outside the immediate sphere that confront parallel challenges, offering broader insights. In healthcare, for instance, the application of NLP for

patient record management demonstrates how technology can overhaul data extraction from unstructured documents. IBM Watson Health's utilization of advanced data analytics exemplifies the transformative potential of such technologies on data accessibility and management.

Synthesis – Informing the Project Plan for the State of Michigan
The examples gleaned from direct and indirect comparators, supplemented by analogous research, inform the project plan by underscoring the necessity of scalable and adaptable technological solutions. These cases have illustrated that machine learning and NLP are capable of substantially improving data management processes. For the Central Procurement Service (CPS), this translates into the deployment of a system that not only resolves current inefficiencies but is also poised to meet future needs. The team will integrate successful elements from each case study into the project plan, crafting a resilient solution for Michigan's procurement data intricacies.

Project Plan



Determine project plan and establish communication

- Assign project scope and specifications

- Create means of communication with client

Develop skills for PDF download, automation, and NLP

- Determine current team skills and skill gaps needed to be filled
- Delegate work and develop skills necessary to complete project requirements

Devise method of PDF download - Requires: Skills developed and research

- Determine best method of download
- Write code for downloading PDF text

Automate PDF download - Requires: Skills developed and client communication

- Communicate with client to determine optimal form of program automation
- Develop automation process
- Implement automation of PDFs with client

Accumulate training data for NLP - Requires: Scraping process coded and automated

- Scrape PDFs for text
- Determine best form of keyword identification
- Create dataset from scraped PDFs

Develop NLP model - Requires: Research, skills, and client communication

- Create model to process text and determine key words
- Integrate keyword NLP into automated processes

Integrate NLP to database functionality - Requires: NLP and download developed and automated

- Coordinate with client to go live with NLP search functionality
- Implement automated download and data scraping
- Train client to utilize new features

Externalities

Due to the nature of the project, the team will encounter few externalities as they already have on-demand access to the data and a clear understanding of the scope and requirements put forth by the client. The majority of potential problems are internal such as an inability to code or develop download and NLP functionality and can be prevented through research and preparation

Priorities and dependencies

Due to the simple nature of the project, priority will be given to the tasks directly related to the two deliverables: an automated PDF downloading program and an automated NLP process to determine key words within text scraped from the PDFs. The timeline above depicts the order in which tasks are to be completed and the prerequisites to accomplishing each task.

Team Roles

1. Data Cleaning Lead: Micah Blackburn

Job: Ensure the quality and integrity of the web-scraped data by cleaning and preprocessing it

Responsibilities:

- Develop a data cleaning and preprocessing pipeline for the collection of the (CPS) dashboard PDF data.
- Handle missing or inconsistent data.
- Remove duplicates and outliers.
- Standardize and format data for NLP analysis.
- Collaborate with the NLP Modeling Lead to refine data preprocessing techniques.

2. NLP Modeling Lead: Jacob Hays

Job: Lead the development and application of NLP models to analyze the PDFs and extract meaningful insights.

Responsibilities:

- Learn about current libraries and technologies to develop NLP pipelines.
- Build and fine-tune NLP models for text extraction and analysis.
- Develop custom algorithms for PDF parsing and content extraction.
- Collaborate with the Data Cleaning Lead to preprocess data effectively.
- Interpret NLP results and generate insights.
- Stay updated on NLP advancements and recommend improvements.

3. Project Manager: Abhinav Tadikonda

Job: Oversee the entire project, ensuring it stays on track, meets deadlines, and achieves objectives.

Responsibilities:

- Develop and manage project plans and timelines.
- Ensure project objectives align with business goals.
- Oversee the execution of the project plan at a high level
- Identify potential obstacles and bottlenecks
- Coordinate resources and schedule for a smooth workflow across major tasks

4. Communications Lead/Client Liaison: Jason Kemp

Job: Act as the primary point of contact between the project team and clients or stakeholders.

Responsibilities:

- Communicate project updates, milestones, and results to clients.
- Address client inquiries and concerns.
- Manage client expectations and feedback.
- Facilitate effective communication within the project team.
- Schedule meetings with the team and the clients.
- Ensure client satisfaction and maintain strong relationships.

5. Editor: Abhinav Tadikonda

Job: Review and enhance the quality and readability of reports, findings, and documentation generated from analysis.

Responsibilities:

- Proofread and edit reports and documents.
- Ensure clarity, grammar, and coherence.
- Verify the accuracy of technical content.
- Collaborate with the NLP Modeling Lead to communicate findings effectively.
- Maintain documentation standards.

6. Lead Presenter: Jason Kemp

Job: Represent the project team in presenting findings and insights to clients or the wider audience.

Responsibilities:

- Prepare and deliver presentations on project progress and results.
- Effectively communicate complex technical information.
- Engage and address questions from the audience.
- Translate technical findings into actionable recommendations.
- Showcase the impact and value of the project.

7. Records Keeper: Micah Blackburn

Job: Maintain organized records of project-related documentation, data, and reports.

Responsibilities:

- Establish a system for storing and cataloging project data.
- Manage version control of documents and reports through Github.
- Facilitate easy retrieval of historical project information and project documents.
- Assist team members in accessing relevant project materials.

To ensure the project's success, a robust strategy for team support and coordination is essential. The roles of Data Cleaning Lead (Micah Blackburn) and NLP Modeling Lead (Jacob Hays) are closely intertwined, which will require regular collaboration. These team members will conduct weekly meetings to align their efforts and ensure that the cleaned data is optimally formatted for NLP analysis. The Project Manager (Abhinav Tadikonda) will oversee this collaboration, providing support and guidance as needed, and ensuring that both roles are aligned with the project's overall objectives.

The Communications Lead/Client Liaison (Jason Kemp) plays a pivotal role in maintaining clear communication lines not only with the client but also within the team. This role will be supported by the Editor (Abhinav Tadikonda), particularly in preparing and reviewing client communications and project reports to ensure accuracy and clarity. Similarly, the Lead Presenter (Jason Kemp) will work closely with the NLP Modeling Lead to translate technical findings into comprehensive presentations.

The Records Keeper (Micah Blackburn) will coordinate with all team members to ensure that all project documents and data are properly stored and easily accessible. This role is crucial for maintaining the project's historical context and aiding in effective information retrieval.

In terms of role adjustments, if the need arises due to workload or unforeseen circumstances, the team is prepared to reallocate tasks among members. For instance, tasks related to data cleaning could be shared between the Data Cleaning Lead and the NLP Modeling Lead if the volume of data increases significantly. Similarly, project management tasks could be distributed between the Project Manager and the Communications Lead to balance the workload effectively. This flexible approach ensures that the team can adapt to changing project demands while maintaining efficiency and meeting deadlines.