

CIRCUITS AND SYSTEMS PERFORMANCE BENCHMARKING USING MACHINE LEARNING

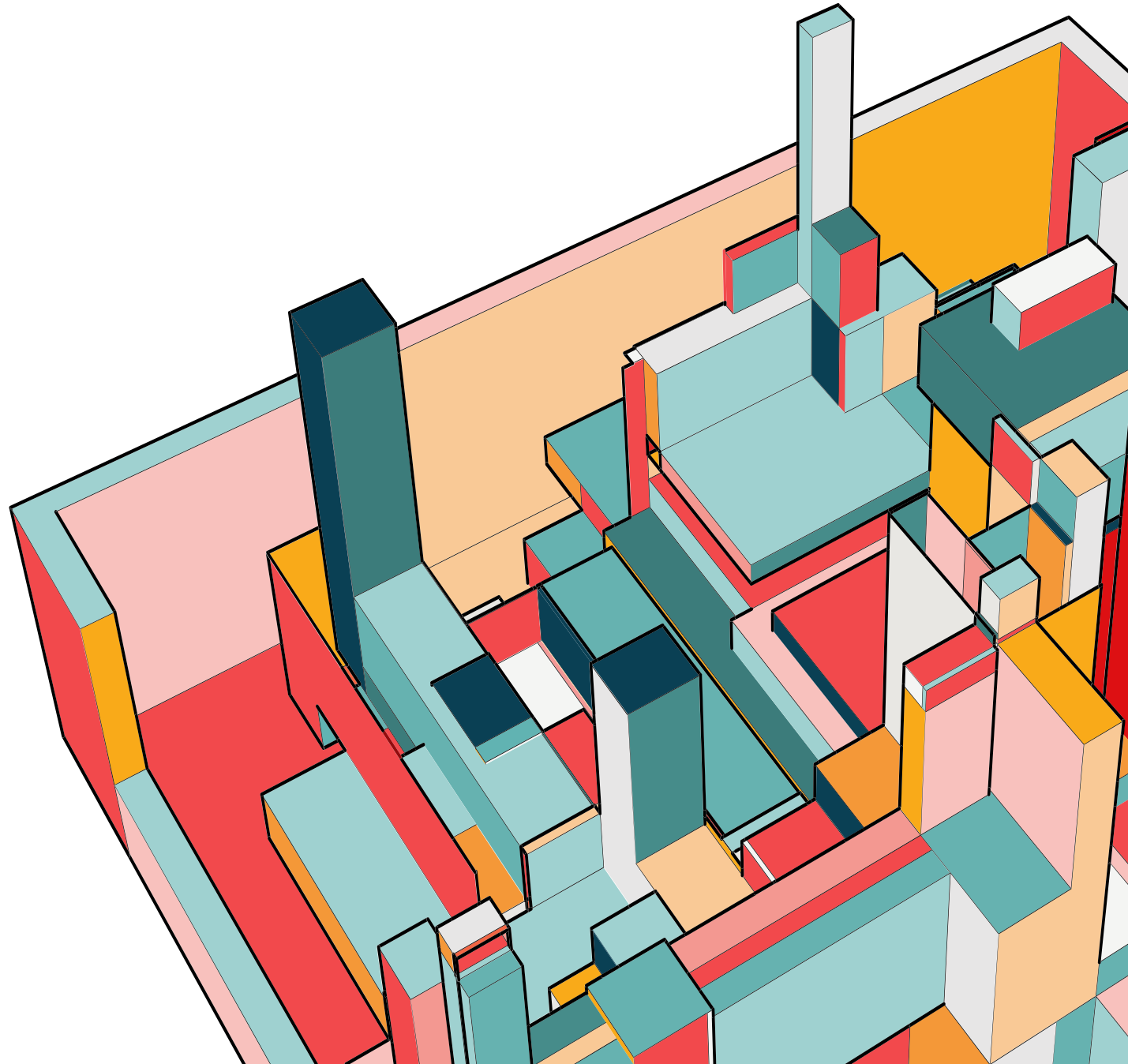
Final Presentation for PURE

Supervisor: Korkut Tokgöz

Members: Deniz Batu, Emir Kantul,
Furkan Akgün

PROBLEM

New electronic devices are being built every day such as power amplifiers. These devices come with documents that present their technical specifications. The problem is that these details are not present in an easy-to-access form in one place. This makes statistical analysis an arduous, labor-intensive task.



PROBLEM

Accessability

To learn a device's specifications, a researcher needs to read the technical documentation of that device. This is time-consuming as one has to go through pages of unnecessary information.

Scale

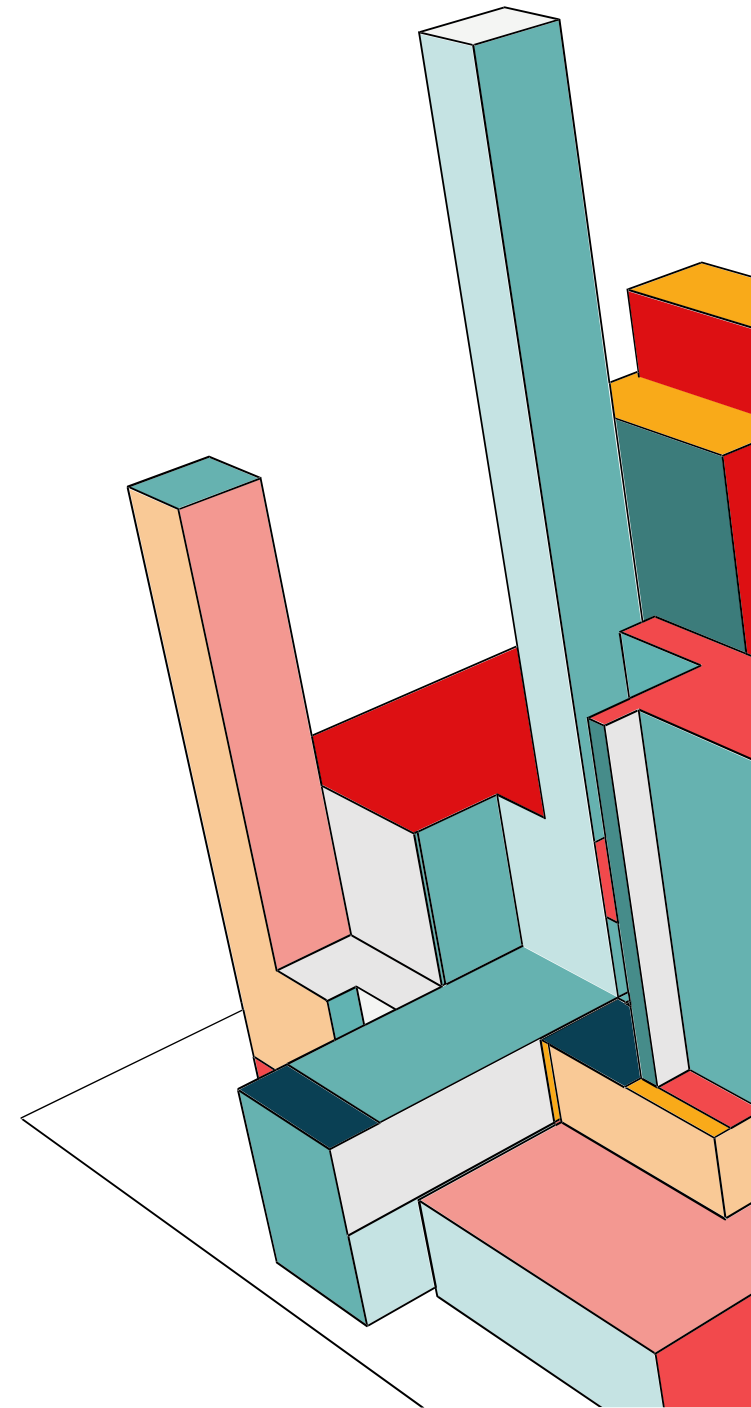
If a researcher wants to see the trend between years and a certain attribute of a certain type of device such as the strength of an MRI machine. The researcher needs to gather information from thousands of papers by hand.

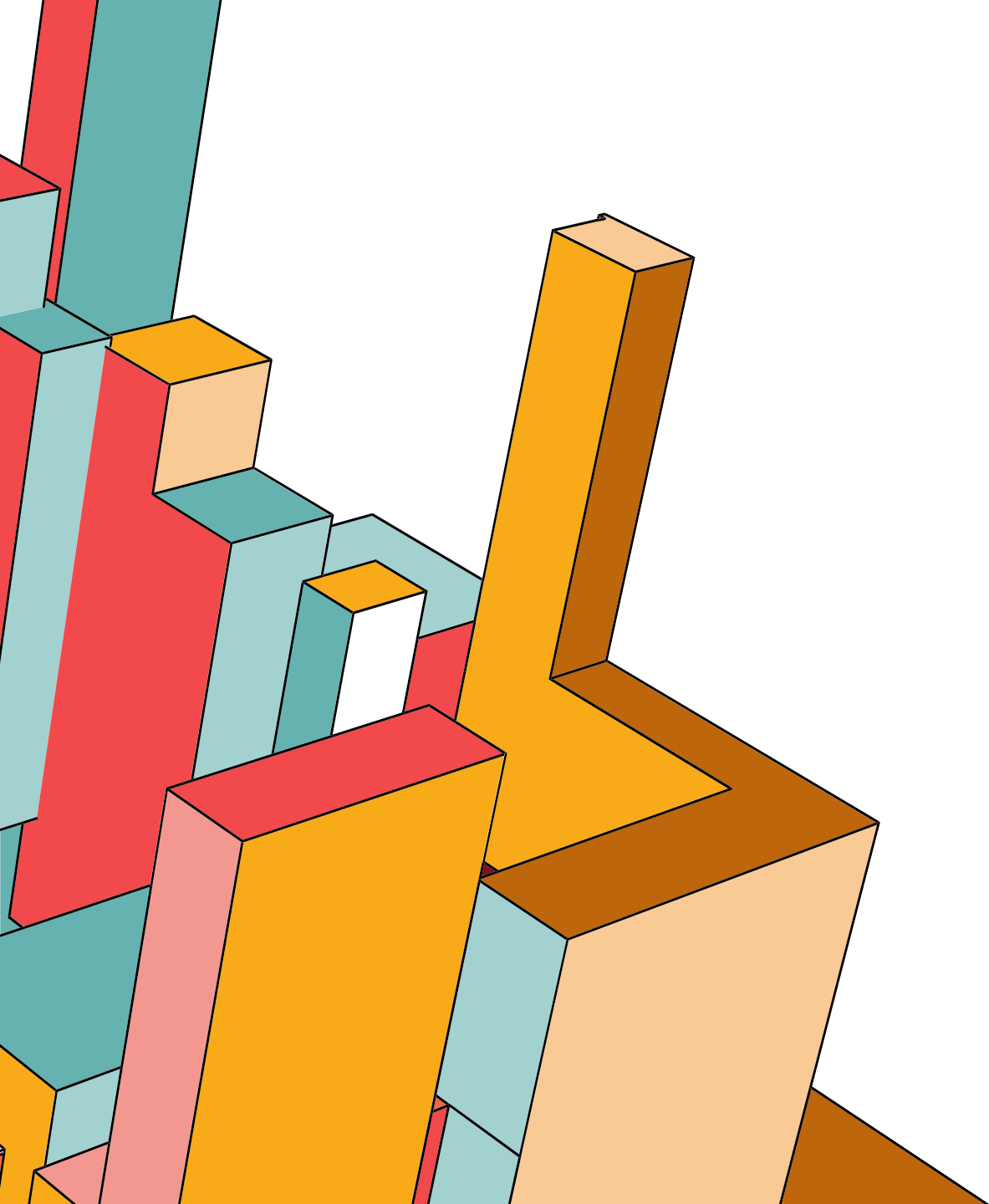
Availability

Most of these specifications are present in research publications. A researcher needs to download a paper belonging to any device of interest.

Ease of Use

All the previously mentioned points make statistical analysis of electronic devices time-consuming and impractical.





SOLUTION

Machine Learning

Machine learning (ML) is a field of inquiry devoted to understanding and building methods that leverage data to improve performance on some set of tasks(Machine Learning, Tom Mitchell, McGraw Hill, 1997)

How

ML can be trained to extract data from research papers as specification listings and sentences have predictable forms. This project aims to leverage ML to achieve this goal and automatize the process of data extraction.

STEPS

List of Papers

A list of research papers belonging to a specific device needs to be present so that the ML can be trained and tested

Paper Format

The papers need to be presented to the ML model in a format that it can process

Model Training

A decision on a training tool needs to be made. The decided model needs to be trained with the papers

Practical Use

The trained model can be used to achieve the goal of this project

PLAN



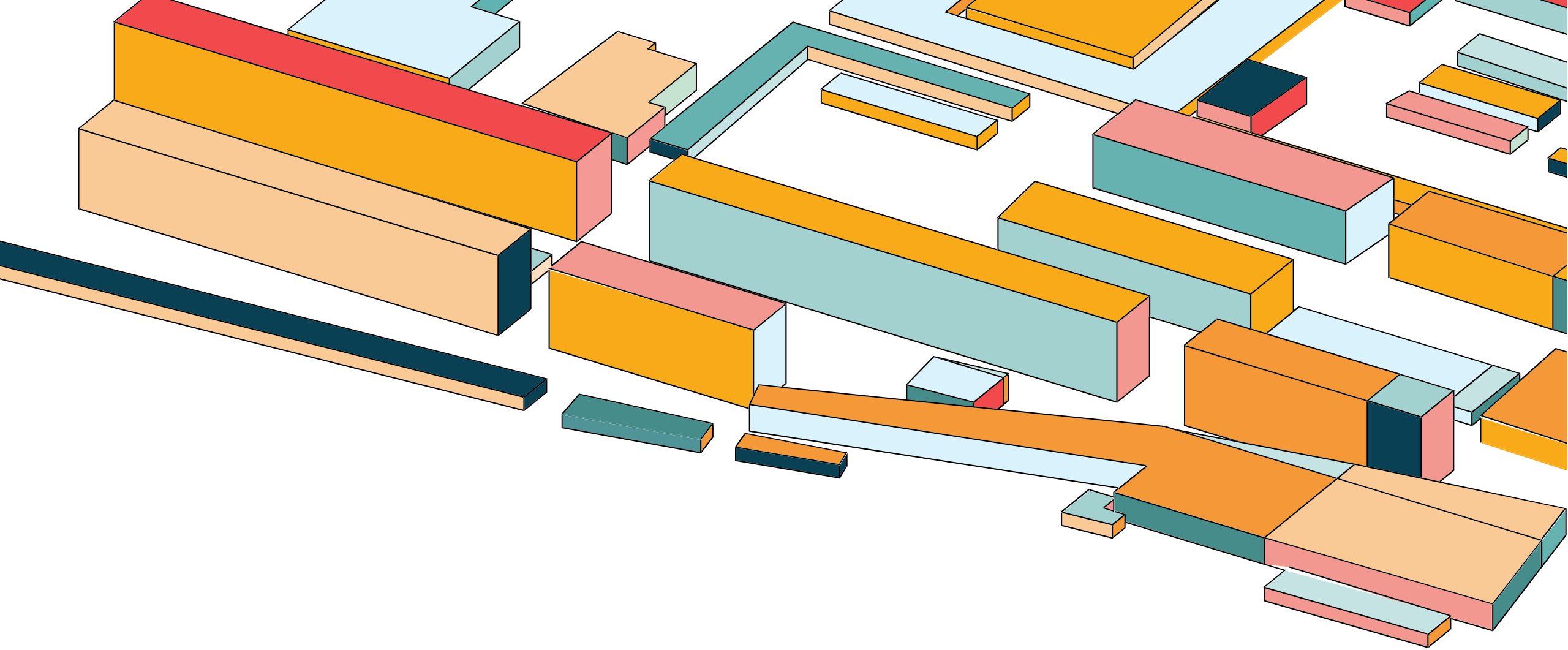
Paper Collection



Paper
Formatization



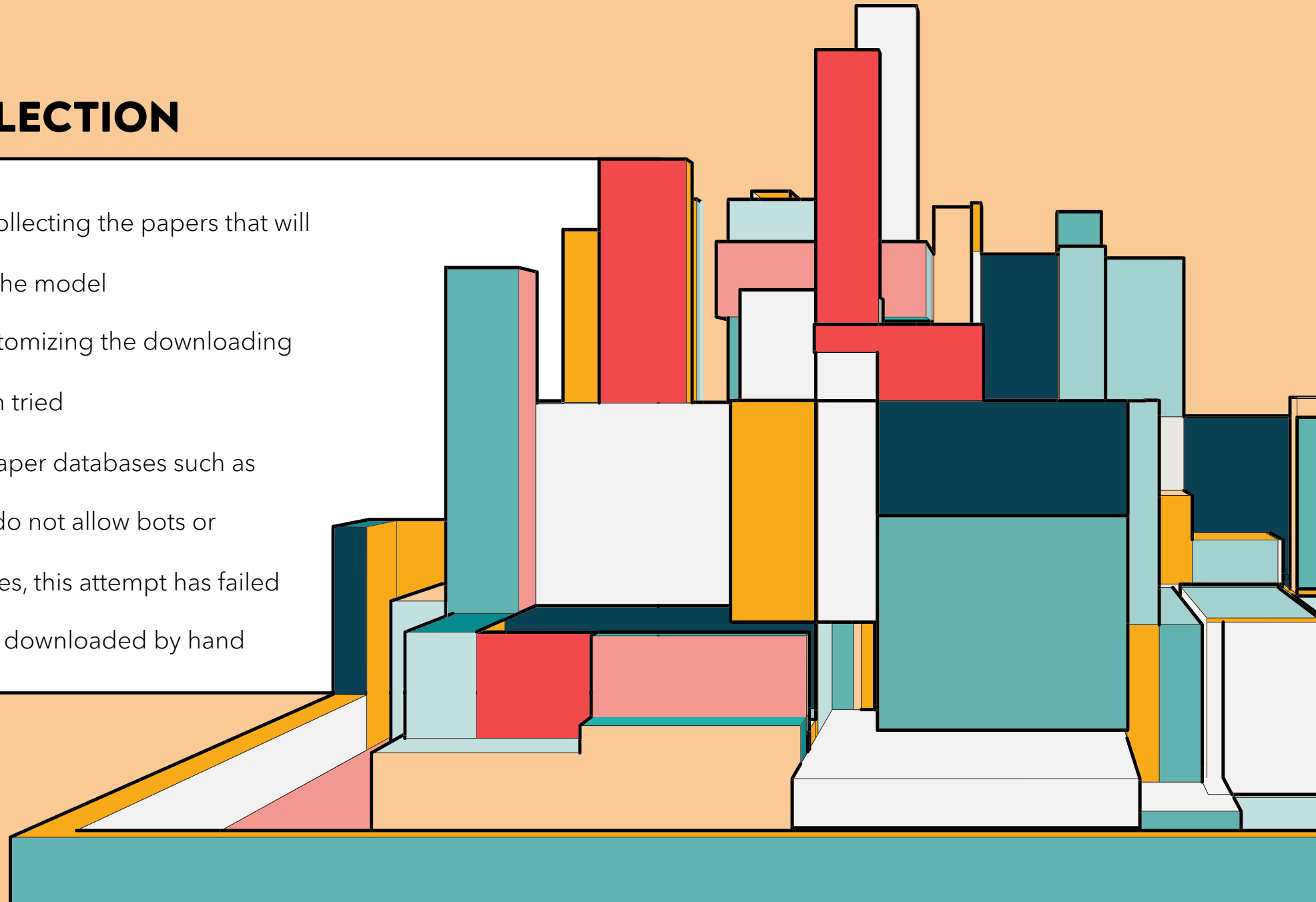
Machine
Learning



PAPER COLLECTION

PAPER COLLECTION

- The first step is collecting the papers that will be used to train the model
- For this effort, automizing the downloading process has been tried
- Since scientific paper databases such as Google Scholar do not allow bots or automated queries, this attempt has failed and papers were downloaded by hand



An abstract graphic on the left side of the slide consisting of several 3D rectangular bars of varying heights and colors. The colors include red, teal, orange, and brown. The bars are arranged in a way that suggests a bar chart or a data visualization, with some bars being taller than others and some overlapping.

Endnote Click as an Alternative Solution

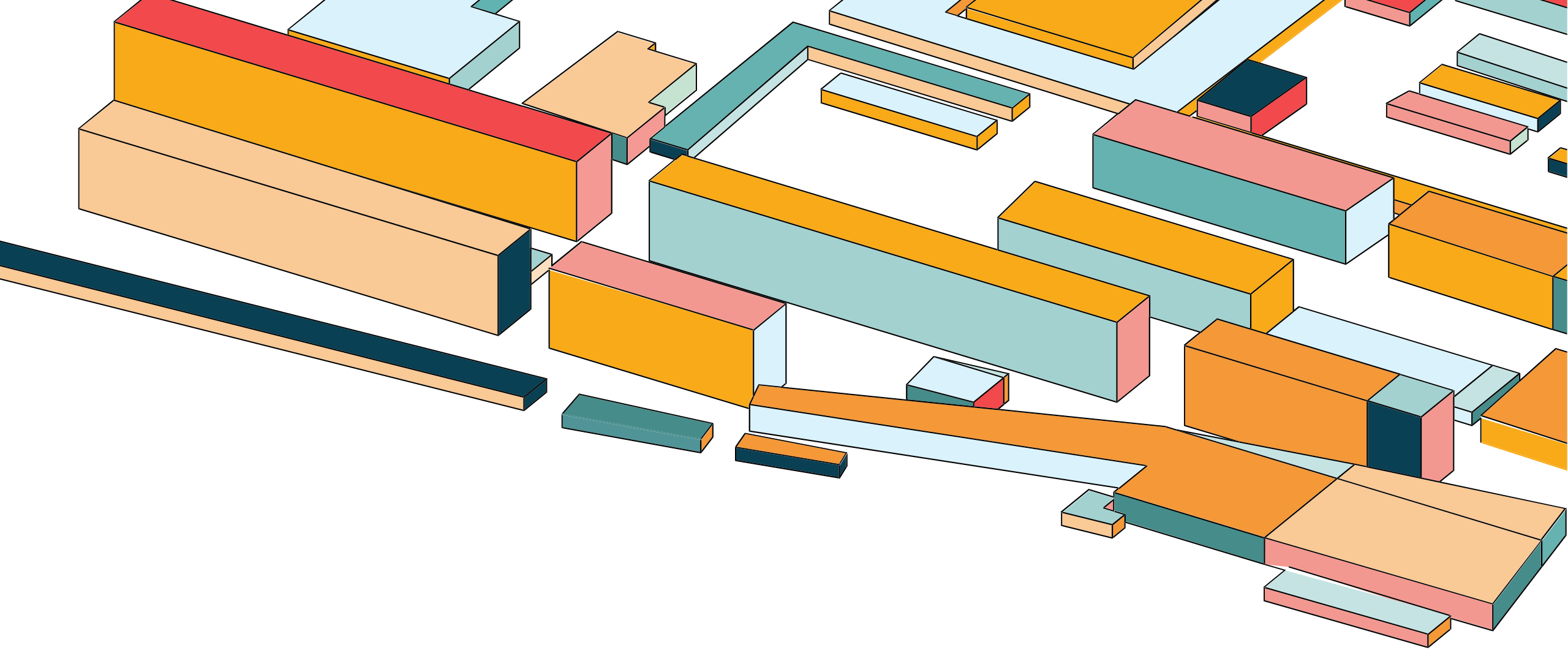
- It is a free application that can be used as a browser plugin that provides significant convenience and compatibility in accessing scientific papers.
- Endnote, which has a higher version and can be installed on the desktop, can also be provided free of charge by many academic institutions.
- Endnote Click provides convenience for creating and updating a personal database for articles on the internet with a single click.
- Processes that require approval therefore prevent automation in every research, such as registration verification processes, can be solved with the Endnote Click application as it will only be done at the login stage.

What is Next?

It has been learned that these databases can be contacted to access requested papers in bulk for scientific purposes

It is aimed to present these authorities with concrete findings of this project to make a case for access authorization

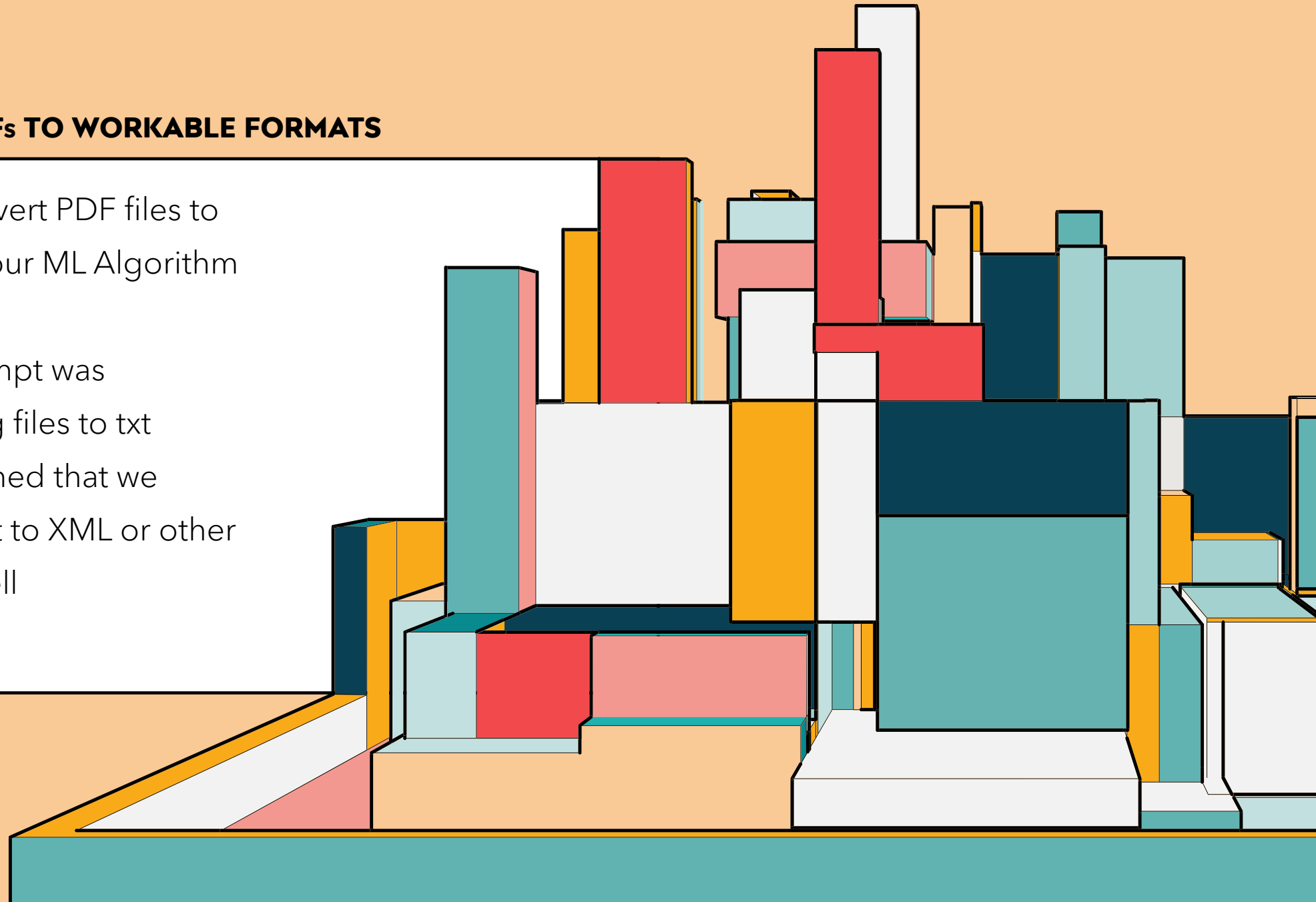
The gathered bulk papers can be processed with automation studios such as Ranorex or Selenium for future work



PAPER FORMATIZATION

CONVERTING PDFs TO WORKABLE FORMATS

- We must convert PDF files to formats that our ML Algorithm can process
- Our first attempt was on converting files to txt
- Later we learned that we could convert to XML or other formats as well



CONVERSION METHODS



PDF
CONVERSION



PDFX



COMPUTER
VISION

Online PDF Conversion Service - HiPDF

Minimum pricing is \$28 for 1000 file conversion per month

Pros

- + No parallel paragraph in one page problems
- + Free API Service up to 200 request
- + Lower price than average
- + Fast conversion

Cons

- Higher prices for conversions more than 200

PDF Conversion Library - Pdfplumber

Written in Python

Pros

- + Easy to use language
- + More option to implement further
- + Very good conversion
- + No parallel paragraph in one page problem

Cons

- Slower

PDFX (XML)

- Regardless of the formatting style used, PDFX is a rule-based system created to reconstruct the logical structure of academic articles in PDF format
- The output of the system is an XML file that relates to geometrical typesetting markers in the original PDF, like paragraph and column breaks, and explains the logical structure of the input article in terms of title, sections, tables, references, etc.

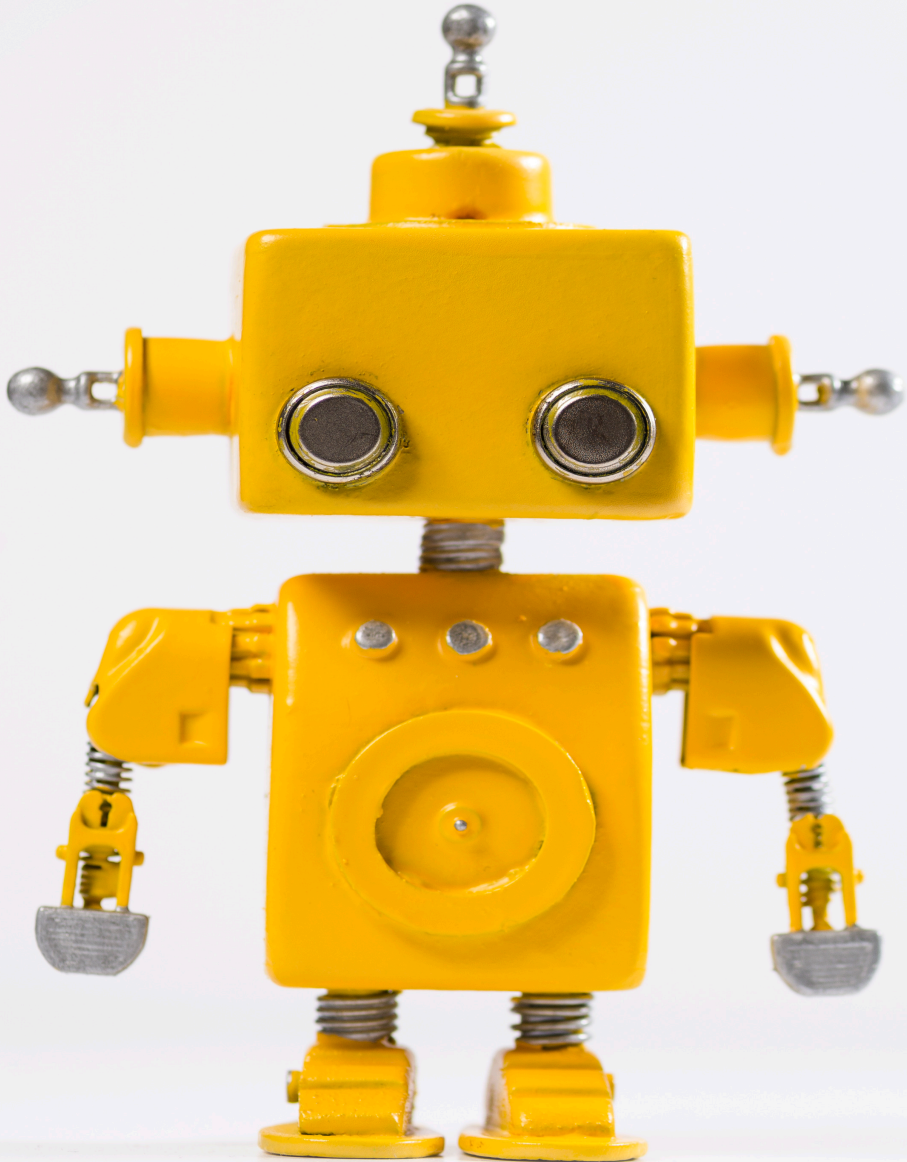
Computer Vision Method

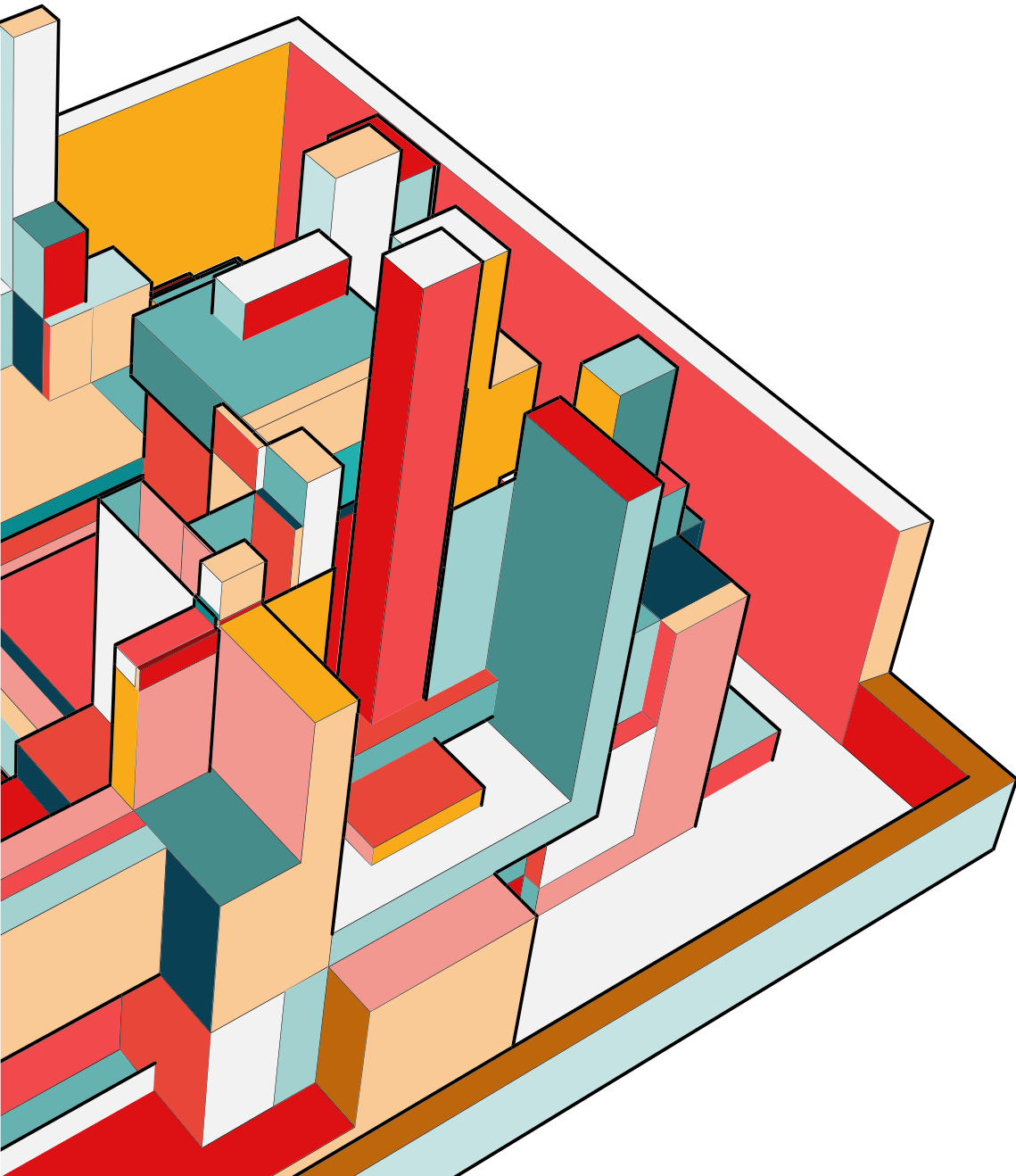
- Optical Character Recognition (OCR) , one of the Computer Vision techniques, can also be used for these purposes as an alternative
- In the OCR, the input is an image that gets converted into the desired characters as an output in a txt format, and OCR can be applied in the Label Studio application

Advantages and Disadvantages

- + In addition to PDF-based documents, OCR can also work on documents in the form of images, handwriting or printed paper
- + Since OCR works on an image basis, it will also be advantageous in examining articles with images
- Performance may be lower than PDF converter applications for simple documents

CHAT GPT-3





FIRST TRIALS OF INFORMATION EXTRACTION USING CHAT GPT-3

- OpenAI introduced ChatGPT (Generative Pre-trained Transformer) as a chatbot in November 2022. It is constructed on top of the GPT-3.5 family of big language models from OpenAI, and it has been enhanced using supervised and reinforcement learning methods.
- ChatGPT is a language prediction model. This indicates that it has a machine learning neural network model that can take text as an input and transform it into the outcome it thinks will be most helpful. This is done by teaching the system to recognize patterns in the massive amount of material on the internet.



what are the attributes of power amplifier on given text:

Abstract - This paper presents an H-bridge class-D power amplifier for digital pulse modulation transmitters. For the proposed class-D amplifier, the drivers of the pull-up and pull-down devices were separated to minimize power loss associated with the shoot-through currents. The amplifier can be driven by binary digital signals generated by a delta-sigma modulator directly. The H-bridge class-D power amplifier system was tested with CDMA IS-95 signals at 800MHz. The drain efficiency of 31% was achieved with output power of 15 dBm and an ACPR of -43 dBc.


I. INTRODUCTION

With the rapid advance of CMOS technology, digital signal processing (DSP) techniques can be used at clock frequencies reaching into the microwave region. This permits the implementation of digital RF systems, implementing functions up to now in the domain of analog circuits.

In digital RF transmitters, signal processing such as base-band signal generation, filtering, and frequency conversion are completed in the digital domain. This digital approach increases the flexibility and programmability of the system, as well as avoiding the problems of aging, variable component values and interconnects difficulties associated with many analog circuits. Fig. 1 shows a possible architecture of digital pulse modulation transmitters [1]. Via DSP techniques, the modulated baseband signals are generated, up-converted, and sent to a band-pass delta-sigma modulator (BPDSM). The BPDSM quantizes the signals into a binary format to drive the following amplifier stage. The associated quantization noise can be spectrally shaped and removed out of band by the feedback loop in the BPDSM. The bandpass filter following the amplifier avoids power dissipation at undesired frequencies to achieve high efficiency. This architecture reduces the system complexity while increasing the flexibility.

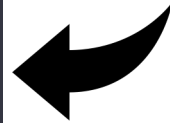
Switching amplifiers are attractive candidates for the digital RF transmitters because of the potential to obtain high system efficiency while driven by digital signals. However, the suitable types of switching amplifiers are limited by the digital driving signals which are non-periodic and broadband. For instance, Class-E amplifiers can operate at RF frequencies efficiently by minimizing the output capacitance loss. However, the zero voltage switching condition for compensating the output capacitance loss cannot be maintained under non-periodic driving conditions, thus the conventional Class-E amplifier cannot achieve high efficiency when driven by the delta-sigma modulated signals.

Voltage-mode Class-D switching amplifiers have the potential to maintain high efficiency when the driving signals are not periodic. However, loss associated with the driving circuits (with limited bandwidth), the active devices (including shoot-through current loss) and filters (poor power recycling) can degrade the performance significantly. Previously, a bandpass delta-sigma Class-S amplifier was demonstrated at 10MHz, showing 3300 drain efficiency with an IM3 of -40dBc [2]. A transformer-coupled amplifier was demonstrated at 170MHz with a drain efficiency of 8% [3]. A Class-D PA with a digital modulator based on quadrature pulse modulation was also demonstrated for EDGE signals

 Regenerate response



Given text is generated with Pdfplumber and fed to chat-gpt with given question.



The attributes of the power amplifier described in the given text are:

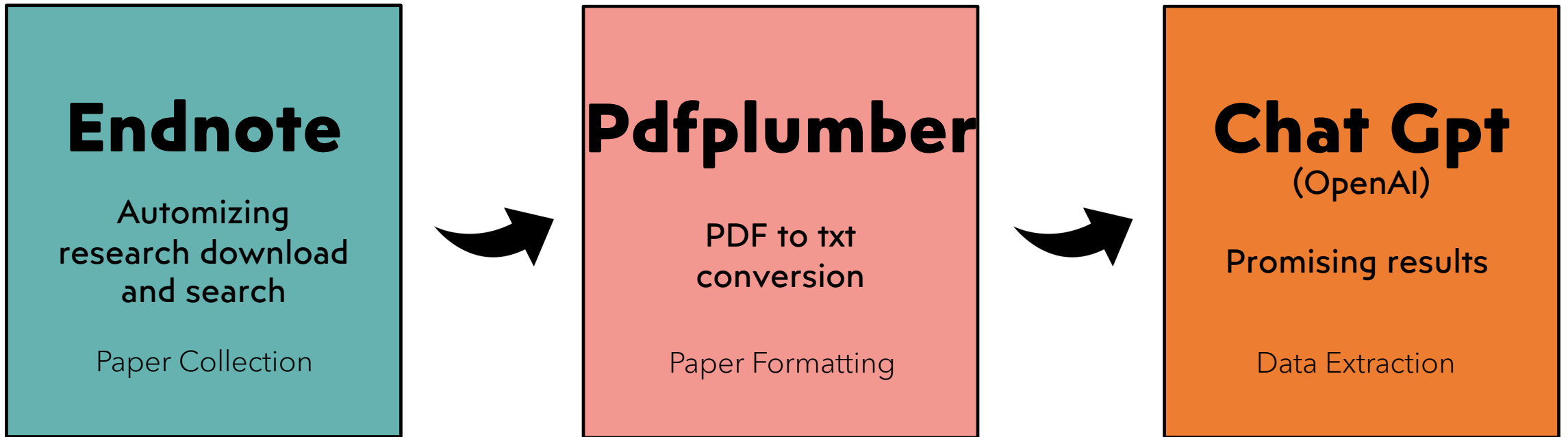


- It is an H-bridge class-D amplifier
- It is designed for use in digital pulse modulation transmitters
- The drivers of the pull-up and pull-down devices are separated to minimize power loss associated with shoot-through currents
- It can be directly driven by binary digital signals generated by a delta-sigma modulator
- It was tested with CDMA IS-95 signals at 800MHz
- It achieved a drain efficiency of 31% with an output power of 15 dBm and an ACPR of -43 dBc
- It is intended for use in digital RF transmitters, where signal processing is done in the digital domain to increase flexibility and programmability, and to avoid the problems associated with analog circuits
- It is a voltage-mode class-D switching amplifier that has the potential to maintain high efficiency when driven by non-periodic signals
- It is intended to compensate for the loss associated with driving circuits, active devices, and filters, which can degrade the performance of the amplifier.



Result with only question. Imagine how many questions we can ask and specialize them for attributes

CONCLUSIONS



FUTURE WORK

Our effort made us find out which methods and tools can be used for our goal of automating the process of extracting statistical information from electric device documents

Endnote is an accessible tool for paper collection which we plan to utilize in the future. There are reliable paid services and open-source approaches for PDF conversions. Artificial Intelligence such as GPT-3 also has shown significant promise.

We are planning to experiment further and start training ML models with the extracted information to make our case for accessing academic databases. Once we can access these databases, we will fully focus on training our models to achieve our goal.