



HEDICALL HISTORY PREDICTION
AND MACHINE LEARNING

MEDICAL
HISTORY

Medical Transcripts GenAI

Windy City Health Coders

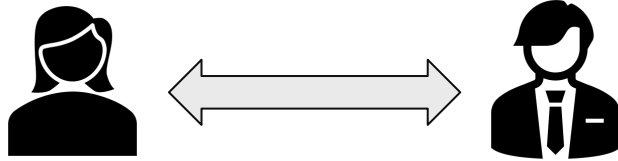
Welly Huang, Kitae Kim

Agenda

- 1 **Business Problem and Purpose**
- 2 **Methodology**
- 3 **Application Build-Up**
- 4 **Next Steps**

1. Business Problem and Purpose

Current System



| Unnamed: 0 | description | medical_specialty | sample_name | transcription | keywords |
|------------|--|----------------------------|---|--|---|
| 0 | A 23-year-old white female presents with comp... | Allergy / Immunology | Allergic rhinitis | SUBJECTIVE: This 23-year-old white female pr... | allergy / immunology, allergic rhinitis, gler... |
| 1 | Consult for laparoscopic gastric bypass. | Bariatrics | Laparoscopic Gastric Bypass Consult - 2 | PAST MEDICAL HISTORY: He has difficulty climb... | bariatrics, laparoscopic gastric bypass, weigh... |
| 2 | Consult for laparoscopic gastric bypass. | Bariatrics | Laparoscopic Gastric Bypass Consult - 1 | HISTORY OF PRESENT ILLNESS: I have seen ABC... | bariatrics, laparoscopic gastric bypass, heart... |
| 3 | 2-D M-Mode, Doppler. | Cardiovascular / Pulmonary | 2-D Echocardiogram - 1 | 2-D M-MODE: , 1. Left atrial enlargement wit... | cardiovascular / pulmonary, 2-d m-mode, dopple... |
| 4 | 2-D Echocardiogram | Cardiovascular / Pulmonary | 2-D Echocardiogram - 2 | 1. The left ventricular cavity size and wall ... | cardiovascular / pulmonary, 2-d, doppler, echo... |
| ... | ... | ... | ... | ... | ... |
| 4994 | Patient having severe sneezes about two to... | Allergy / Immunology | Chronic Sinusitis | HISTORY: I had the pleasure of meeting and ... | NaN |
| 4995 | This is a 14-month-old baby boy Caucasian who... | Allergy / Immunology | Kawasaki Disease - Discharge Summary | ADMITTING DIAGNOSIS: Kawasaki disease, DISCH... | allergy / immunology, mucous membranes, cerym... |
| 4996 | A female for a complete physical and follow u... | Allergy / Immunology | Followup on Asthma | SUBJECTIVE: This is a 42-year-old white fema... | NaN |
| 4997 | Mother states he has been wheezing and coughing. | Allergy / Immunology | Asthma in a 5-year-old | CHIEF COMPLAINT: This 5-year-old male presen... | NaN |
| 4998 | Acute allergic reaction, etiology uncertain... | Allergy / Immunology | Allergy Evaluation Consult | HISTORY: A 34-year-old male presents today a... | NaN |

There is no dashboard for patients record.

It is hard to check the status of the patients

Piloted Recommendation Engine



Doctor

Inputting the symptoms and transcription

Dashboard Trend of Data



Patients

Inputting the symptoms

Getting predicted disease and the possible remedies

1. Business Problem and Purpose

Current Problem

It is currently Hard to...

Lacking a dashboard for patient records makes it difficult to check the status of the disease trend of patients.

Inconsistent data extraction from various medical document formats leads to unreliable medical insights.

Poor integration with existing healthcare systems results in data silos and reduced workflow efficiency.

Delays in data extraction and analysis hinder timely clinical decision-making.



Model Benefits

Real-time visualization of patient records allows healthcare providers to **quickly assess** patient status and deliver appropriate care.

Standardized data extraction processes ensure consistent and reliable medical insights, enhancing decision-making.

Seamless **integration with EHR systems** improves workflow efficiency and provides comprehensive access to patient information.

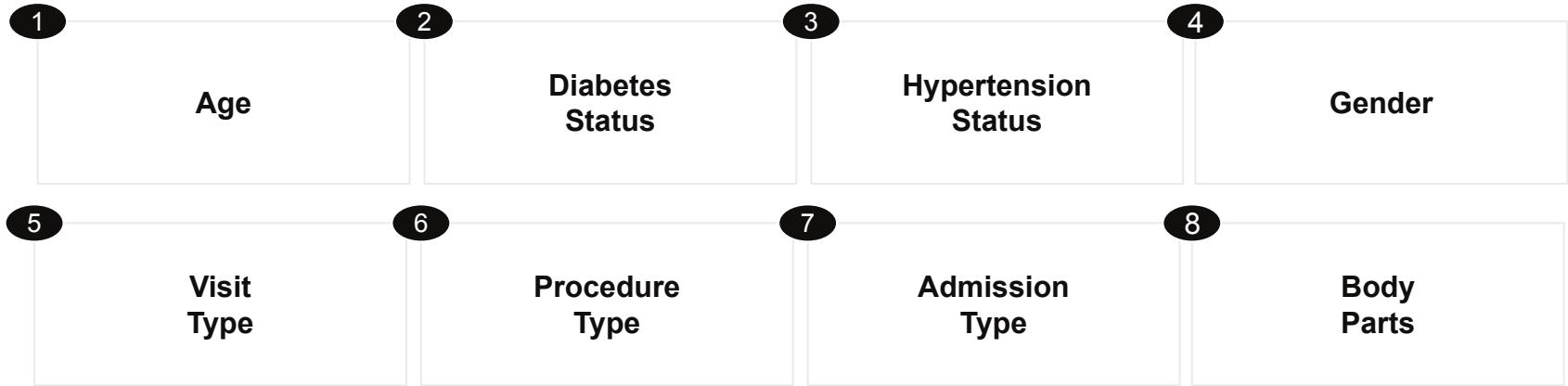
Real-time data processing enables **immediate clinical decisions**, improving patient care and outcomes.

2. Methodology: Data Transition

EHR systems record words like **Description, Medical Specialty, Transcription** transition



Our Medical Dictionary (ICD by WHO) tries to extract key factors from medical history words data



2. Methodology: Data Transition

Results

Medical History

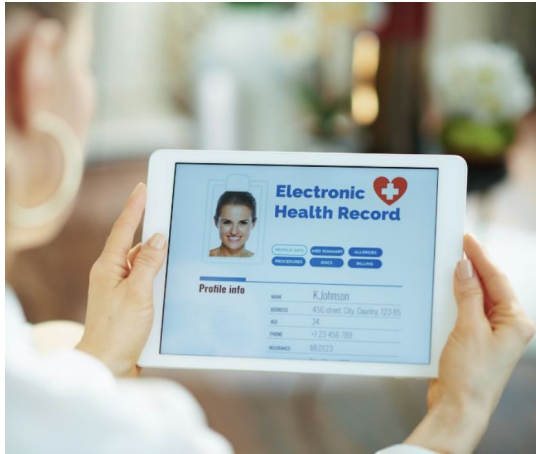
| Unnamed: 0 | description | medical_specialty | sample_name | transcription | keywords |
|------------|--|----------------------------|---|---|---|
| 0 | A 23-year-old white female presents with comp... | Allergy / Immunology | Allergic Rhinitis | SUBJECTIVE:, This 23-year-old white female pr... | allergy / immunology, allergic rhinitis, aller... |
| 1 | Consult for laparoscopic gastric bypass. | Bariatrics | Laparoscopic Gastric Bypass Consult - 2 | PAST MEDICAL HISTORY:, He has difficulty climb... | bariatrics, laparoscopic gastric bypass, weigh... |
| 2 | Consult for laparoscopic gastric bypass. | Bariatrics | Laparoscopic Gastric Bypass Consult - 1 | HISTORY OF PRESENT ILLNESS:, I have seen ABC ... | bariatrics, laparoscopic gastric bypass, heart... |
| 3 | 2-D M-Mode. Doppler. | Cardiovascular / Pulmonary | 2-D Echocardiogram - 1 | 2-D M-MODE:, 1 Left atrial enlargement wit... | cardiovascular / pulmonary, 2-d m-mode, dopple... |
| 4 | 2-D Echocardiogram | Cardiovascular / Pulmonary | 2-D Echocardiogram - 2 | 1. The left ventricular cavity size and wall ... | cardiovascular / pulmonary, 2-d, doppler, echo... |
| ... | ... | ... | ... | ... | ... |
| 4994 | Patient having severe sinusitis about two to ... | Allergy / Immunology | Chronic Sinusitis | HISTORY:, I had the pleasure of meeting and s... | NaN |
| 4995 | This is a 14-month-old baby boy Caucasian who... | Allergy / Immunology | Kawasaki Disease - Discharge Summary | ADMITTING DIAGNOSIS:, Kawasaki disease, DISCH... | allergy / immunology, mucous membranes, conjun... |
| 4996 | A female for a complete physical and follow u... | Allergy / Immunology | Followup on Asthma | SUBJECTIVE:, This is a 42-year-old white fema... | NaN |
| 4997 | Mother states he has been wheezing and coughing. | Allergy / Immunology | Asthma in a 5-year-old | CHIEF COMPLAINT:, This 5-year-old male present... | NaN |
| 4998 | Acute allergic reaction, etiology uncertain, ... | Allergy / Immunology | Allergy Evaluation Consult | HISTORY:, A 34-year-old male presents today s... | NaN |

NLP transition

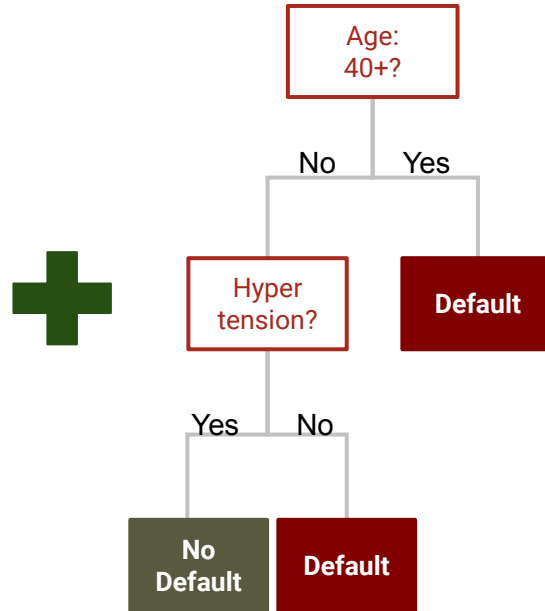
| keywords | age | diabetes | hypertension | heart_disease_type | has_heart_disease | gender | visit_type | procedure_type | admission_type | body_parts | symptoms | diagnosis |
|---|-----------|----------|--------------|--------------------|-------------------|--------|------------|----------------|----------------|---|---|---|
| allergy / immunology, allergic rhinitis, aller... | 23.000000 | 0 | 0 | Other | 0 | Male | Emergency | Other | Other | ear, l/rg, neck, throat | ear drainage, dysphoria | allergy, asthma |
| bariatrics, laparoscopic gastric bypass, weigh... | 45.842075 | 1 | 1 | heart disease | 1 | Female | Other | Consultation | Other | ankle, artery, back, bladder, chest, ear, foot... | ankle pain, arrhythmia, back pain, bleeding, c... | allergy, arthritis, asthma, cancer, chronic ob... |
| bariatrics, laparoscopic gastric bypass, heart... | 42.000000 | 1 | 1 | heart disease | 1 | Female | Other | Consultation | Other | abdomen, ankle, artery, back, arrhythmia, back, bladder, chest... | ankle pain, arrhythmia, back pain, bleeding, c... | arthritis, asthma, cancer, chronic obstructive... |
| cardiovascular / pulmonary, 2-d m-mode, dopple... | 45.842075 | 0 | 0 | cardiovascular | 1 | Male | Emergency | Other | Other | Other | Other | heart disease |
| cardiovascular / pulmonary, 2-d, doppler, echo... | 45.842075 | 0 | 1 | cardiovascular | 1 | Male | Other | Other | Other | artery, ear | Other | heart disease, hypertension |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| bariatrics, laparoscopic gastric bypass, gast... | 27.000000 | 0 | 0 | Other | 0 | Female | Emergency | Other | Inpatient | call, chest, ear | abdominal pain, call pain, chest pain, hoarse... | multiple sclerosis, obesity |
| bariatrics, kerry crag, medlet, nrlsystem... | 65.000000 | 1 | 1 | cardiac | 1 | Female | Emergency | Surgery | Other | abdomen, ankle, brain, chest, ear, eye, heart... | ankle swelling, bleeding, breath sound, chest... | arthritis, cancer, carpalis, depression, diab... |
| bariatrics, elective surgical weight loss, ear... | 51.000000 | 1 | 1 | Other | 0 | Female | Other | Other | Other | abdomen, back, breast, chest, ear, knee, muscle... | back pain, edema, weight loss | arthritis, breast cancer, diabetes, gastroes... |
| allergy / immunology, chronic gastric, warr... | 65.000000 | 0 | 0 | Other | 0 | Male | Emergency | Other | Other | ear, rib, thyroid | depression, dysphagia, fatigue, irritability... | allergy, asthma, depression, fibromyalgia, gas... |
| allergy / immunology, mucous membranes, conjun... | 45.842075 | 0 | 0 | cardiac | 1 | Male | Emergency | Other | Other | Other | edema, fever, rash, redness | allergy, arthritis, conjunctivitis, heart dis... |

2. Methodology : Disease Prediction

EHR Medical History



Random Forest Tree Models

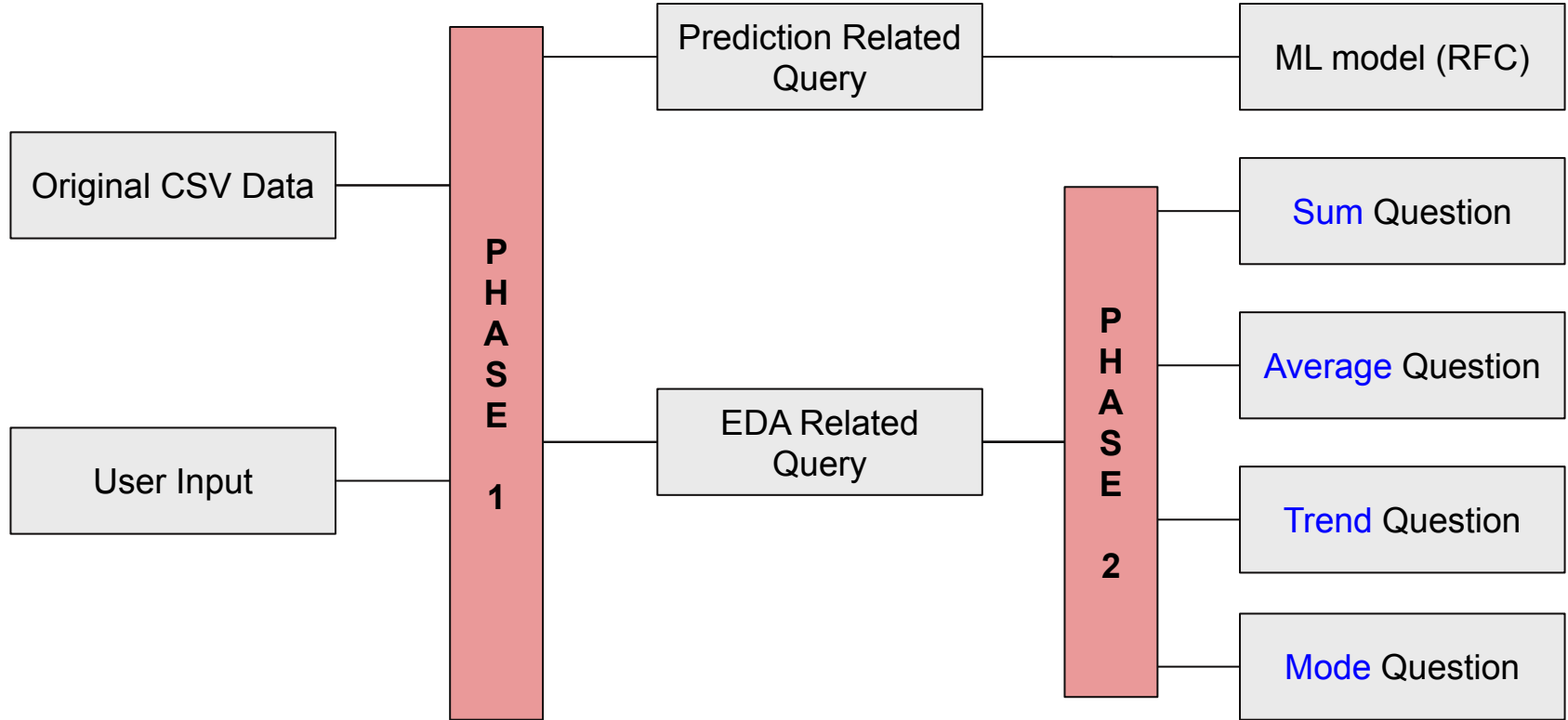


- OneHotEncoder
- RandomForest
- MultiLabelBinarizer
- F1 score: 80%
- Precision: 96%

Disease Prediction:

asthma, allergy, arthritis, asthma,
cancer, chronic ob...
multiple sclerosis, obesity

3. Application: Modeling Framework



3. Application: Solutions for Question 1 - 6

Q1: How many patients have been diagnosed with diabetes in the dataset?

Q2: What is the average age of the patients in the dataset?

Q3: Identify the most common medical condition mentioned in the patient records. Provide the count of occurrences.

Q4: How many patients have been prescribed medication for hypertension?

Q5: Perform a trend analysis on the incidence of heart disease over the years covered in the dataset.
Summarize your findings.

Q6: Can we predict the likelihood of a patient being diagnosed with diabetes based on their medical history?

Link: <https://windy-city-health-coders-medical-transcripts-genai.streamlit.app/>

3. Application: Challenge - 2

- Develop a language model that accurately summarizes the diagnosis or treatment plan for a patient based on their medical history and symptoms distributed in indexed data.
- Develop a language model that accurately answers real-time questions within information from medical records, such as patient demographics, medical history, and examination findings?
- Develop a language model that accurately generates some sections of medical reports and presents the information in a more accessible format for patients or healthcare providers.
- Develop an AI model that can identify potential drug interactions or adverse reactions based on a patient's medical history and current medications?
- Develop an AI model that can predict the likelihood of a patient developing a particular medical condition based on their medical history and demographic information.

3. Application: Let us try predicting with user input symptoms / medical history...

If a patient input the symptoms they have gotten...

“ I am Domingo Green, 45-years-old. Recently, I've been experiencing symptoms related to my cardiovascular health, including irregular heartbeats, chest pain, and shortness of breath. There have been frightening moments..... ”

3. Application: Very accurate...!

Windy City Health Coders - Medical Transcripts GenAI

Upload a CSV file



Drag and drop file here

Limit 200MB per file • CSV

Browse files



mtsamples_with_rand_names.csv 16.3MB



Ask me anything about this dataset: (if you need disease prediction, type "please predict".)

please predict

The query 'please predict' contains a prediction-related keyword.

This is the disease prediction model.

Please enter your symptoms and medical history here.

I am Domingo Green, 45-years-old. Recently, I've been experiencing symptoms related to my cardiovascular health, including irregular heartbeats, chest pain, and shortness of breath. There have been frightening moments where I felt like I was having a heart attack, with severe ches

Your predicted diseases are: coronary artery disease; heart disease

4. Further Steps

Dashboard Development

- Treatment Options and Plan
- Lifestyle and Management Tips
- Preventive Measures
- Additional Tests and Diagnostics:
- Rehabilitation Programs

Like the left, Our dashboard will be include next plans and remedy ways which are easy for Users to follow-up

Model Optimization

“ I have a left lateral malleolus fracture and I am 25-year-old.”

Your predicted diseases are: Other

If we put this short sentence by a patient, it is classified for more health check-up needed code(“Other”). So, we will collect the data and develop and optimize the model with technics

The End.

Thank you for listening!

Questions?

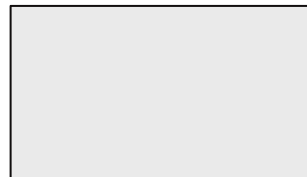
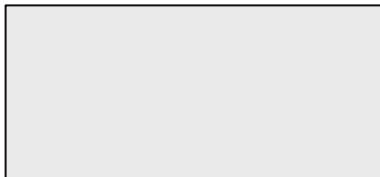
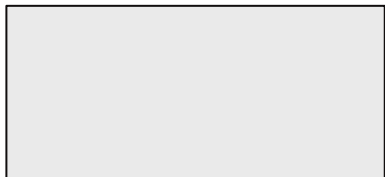
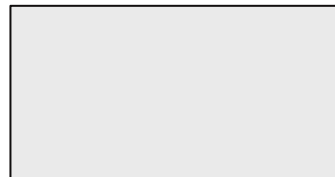
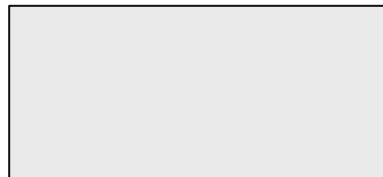
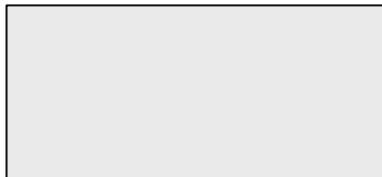
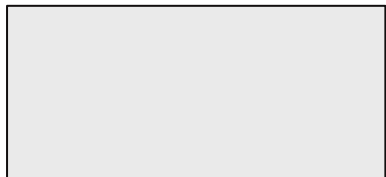
Questions?

d

EDA

2. Prediction and Guid (Doctor and Patients)

Putting Symptoms Prediction and Guid (Doctor and Patients)



Identify the primary objectives and purpose.

- Support the purpose with relevant context.
- Describe the data set(s) including imitations, unit of analysis, time window for feature, model engineering, validation, and development samples (as needed)

Methodology

- Describe the tools used to understand the relevant structure in data.
- Provide a clear description of modeling framework including:
 - Model selection justification
 - Model descriptions
 - Feature engineering/transformations
 - Model engineering /development
 - Validation methodology

Results

Address research objectives and provide:

- Presentation synthesis
- Logical Assumptions
- Descriptive analytics
- Visual storytelling

Modeling Results:

- Model Performance
- Validation Results (you can quantify the performance on below provided sample questions)

Conclusions and Recommendations:

What to Build

Develop a Gen AI application using foundational models (LLM & MLLM) to enhance the analysis of unstructured (text & image), semi-structured (i.e. CSV), and structured data (i.e. tables), such as research papers, financial and medical mixed data documents. This will help organizations improve the quality of their analysis, boost productivity, and streamline workflows. Participants' submissions will be evaluated using sample data provided on Canvas for each use case category.

Challenge 2 (healthcare):

1. **Question:** How many patients have been diagnosed with diabetes in the dataset?
2. **Question:** What is the average age of the patients in the dataset?
3. **Question:** Identify the most common medical condition mentioned in the patient records. Provide the count of occurrences.
4. **Question:** How many patients have been prescribed medication for hypertension?
5. **Question:** Perform a trend analysis on the incidence of heart disease over the years covered in the dataset. Summarize your findings.
6. **Question:** Can we predict the likelihood of a patient being diagnosed with diabetes based on their medical history?

MORE SAMPLE BUSINESS QUESTIONS (for further testing)

Challenge-2:

- Develop a language model that accurately summarizes the diagnosis or treatment plan for a patient based on their medical history and symptoms distributed in indexed data.
- Develop a language model that accurately answers real-time questions within information from medical records, such as patient demographics, medical history, and examination findings?
- Develop a language model that accurately generates some sections of medical reports and presents the information in a more accessible format for patients or healthcare providers.
- Develop an AI model that can identify potential drug interactions or adverse reactions based on a patient's medical history and current medications?
- Develop an AI model that can predict the likelihood of a patient developing a particular medical condition based on their medical history and demographic information.

Below are templates

Agenda

1

dd

2

dd

3

dd
d

4

d

5



dd

6

dd

ddd

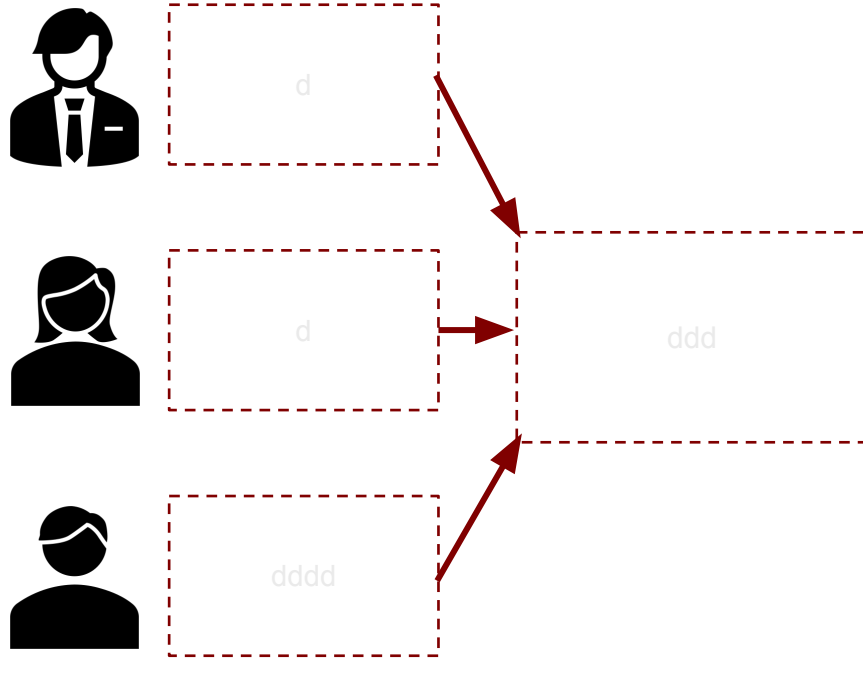
dddd

| Data Field | Description | Received | Notes |
|------------|-------------|---|-----------------------------------|
| dd | dd |  | ~1.4% of transactions are missing |
| dddd | dd |  | |
| dd | dd | <input type="checkbox"/> | |
| dd | dd | <input type="checkbox"/> | |
| dd | ddd | <input type="checkbox"/> | ddd |

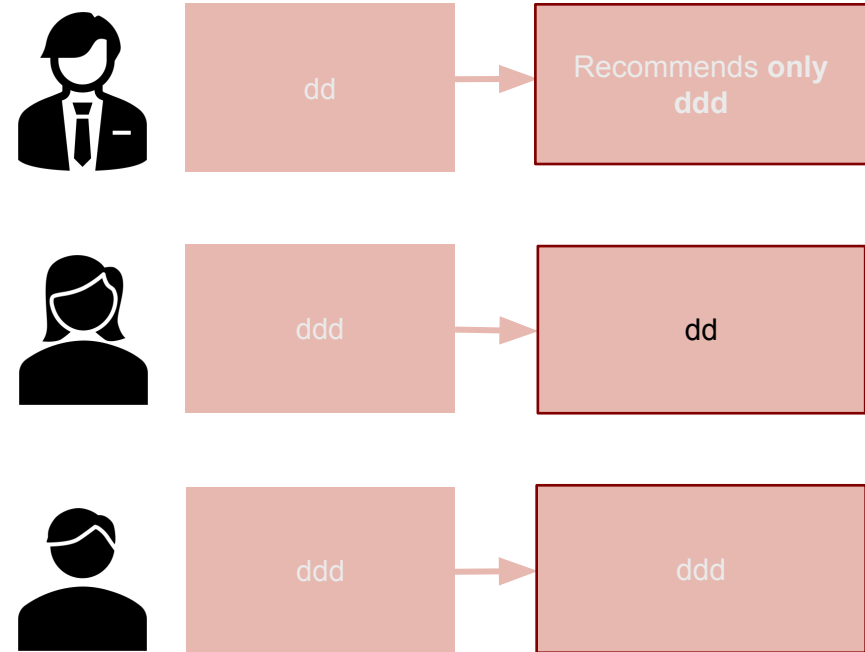
Additional data requests to McDonald's team

dddd

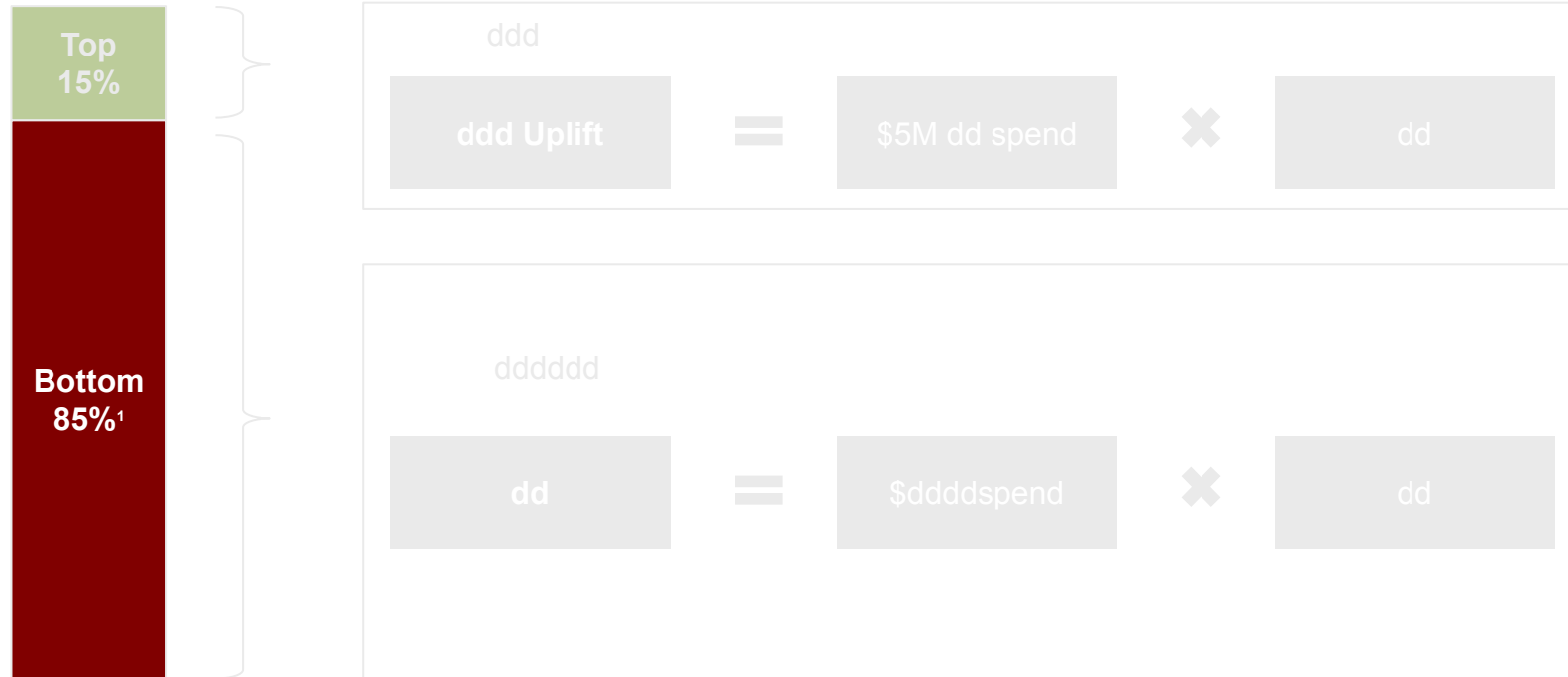
Current System



Piloted Recommendation Engine



Pilot Results – Top 15% generates significant uplift; Bottom 85% revenues can be increased



dddd

dddd
ddd



Identified Opportunity: Bottom 85% of customers have not recently visited the website

RFM Customer Segments

| | Customer Segment | Recency Score | Pilot Revenue Uplift |
|-------------------------|----------------------|---------------|----------------------|
| Top 15% | Big Spenders | High | 20% |
| | Regular Shoppers | | |
| Bottom 85% ¹ | Casual Shoppers | Low | 3-5% |
| | Former Regulars | | |
| | Occasional Splurgers | | |
| | Window Shoppers | | |

Assumption: low recency indicates no interaction w/ recommendation engine

Unlocking Growth: Re-energizing bottom 85% with targeted marketing offers

Bottom 85%¹ customer segments: receive e-mail offers containing products recommended by new engine

Vary offers based on customer segment characteristics

1

Casual Shoppers
Reliable repeaters,
high-value customers

2

Former Regulars
Many purchases, fading
interest

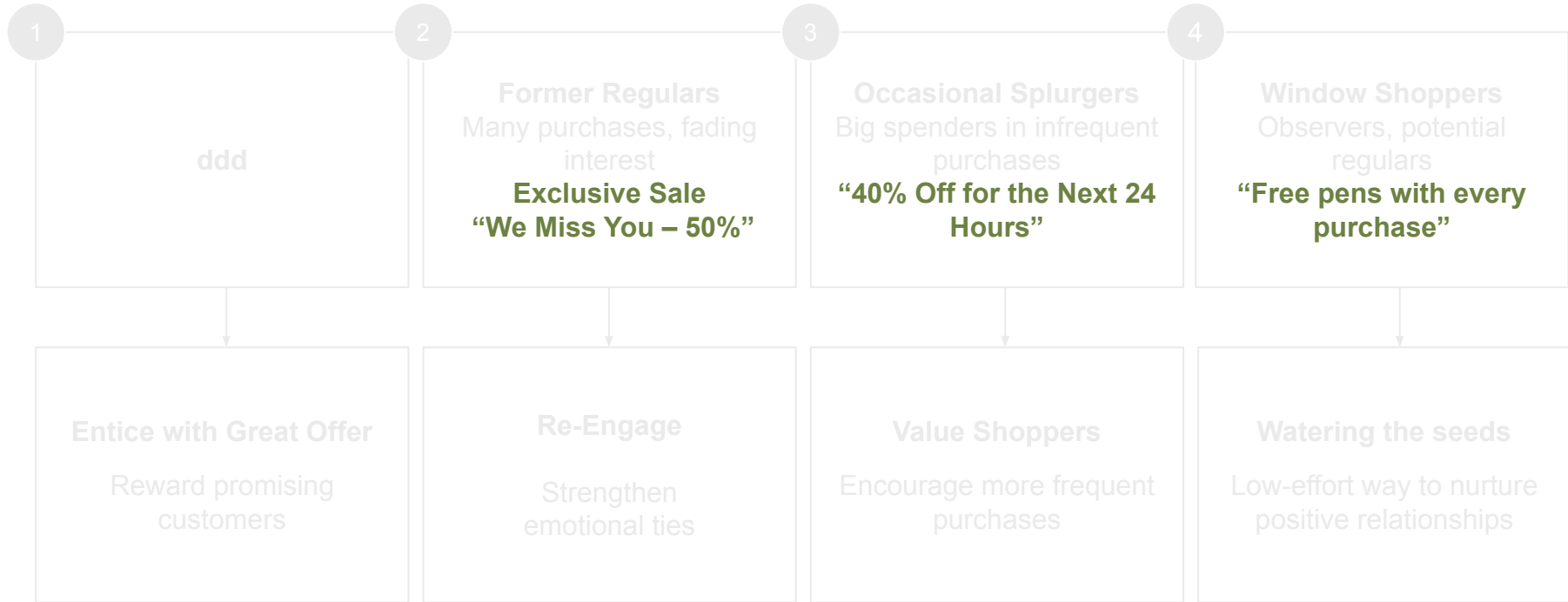
3

Occasional Splurgers
Big spenders in infrequent
purchases

4

Window Shoppers
Observers, potential
regulars

Targeted Marketing Strategies: Capitalizing on segment characteristics



Calculating Revenue Uplift from Marketing: Exploring the case of a single Former Regular

Revenue Uplift Methodology:

$$\text{Revenue Uplift} = \text{Current Revenue per Customer} \times \% \text{ Revenue Uplift} \times \% \text{ Conversion Rate}$$

Example: What happens to a single “Former Regular”?

$$\text{Revenue Uplift for a single "Former Regular"} = \sim \$71 \times 40\%^1 \text{ Revenue Uplift} \times 50\%^2 \text{ Conversion Rate}$$

Sense Check:
40% revenue uplift means buying another book at ~\$28

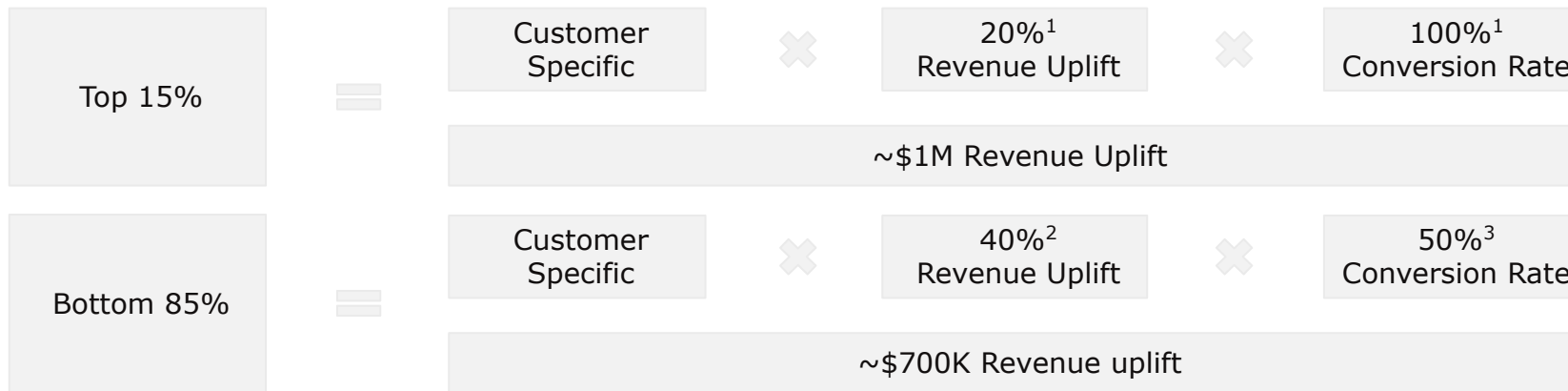
~\$14 Revenue Uplift

Overall Recommendation is expected to create \$1.7M in revenue uplift

Revenue Uplift Methodology:



Extrapolated to Entire Population:



Problem Definition: Poor default classification leads to heavy costs

Current Problem

High default rates on approved device plans; current rate at **11.5%** (115K customers)

Current revenue identification strategy declines large volume of safe customers (>100K customers)

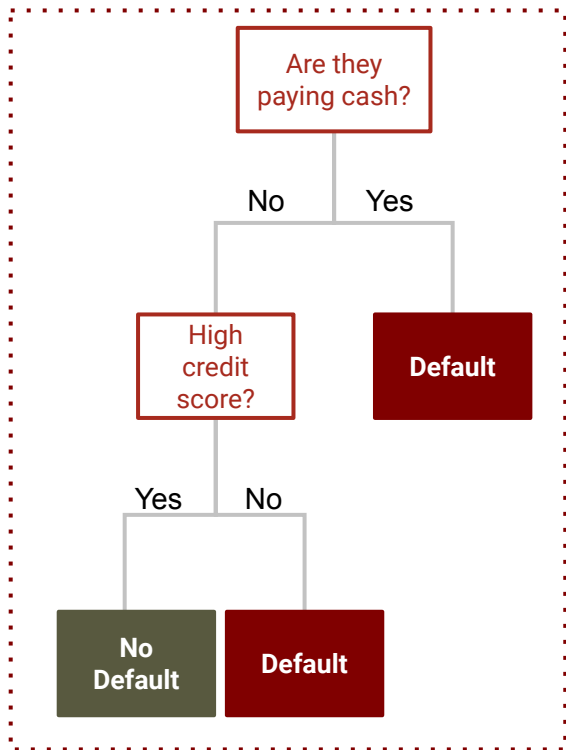
Model Benefits

Reduce Defaults:
Decline contracts to high-risk customers

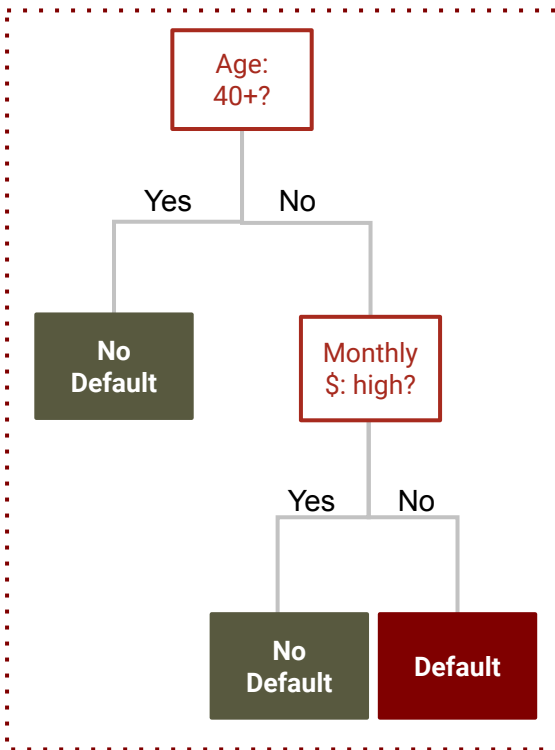
Increase Customer Base:
Offer more contracts to safe customers

Random Forest Model: A Sum of Randomly Generated Decision Trees

Example Decision Tree 1



Example Decision Tree 2



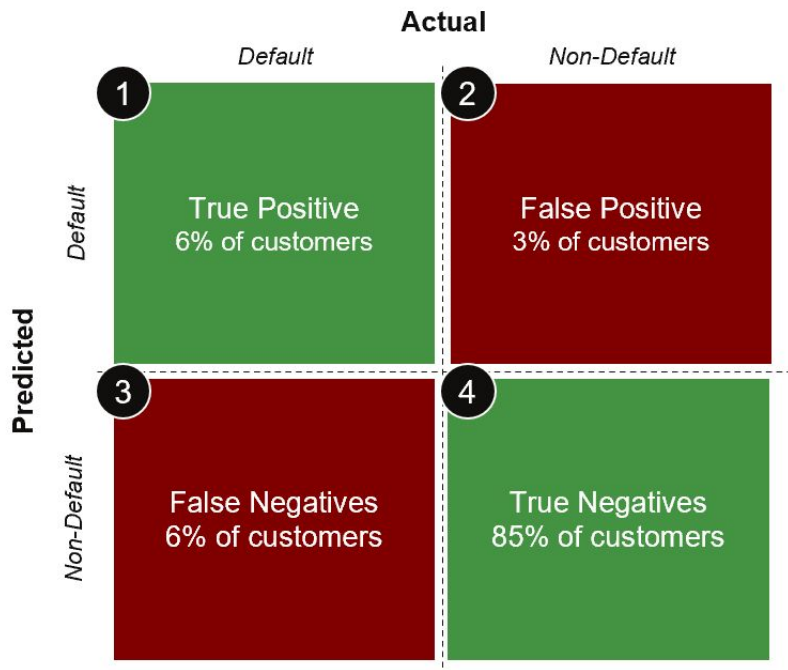
...



Model Performance

Confusion Matrix

Random Forest Model, n= 9,934

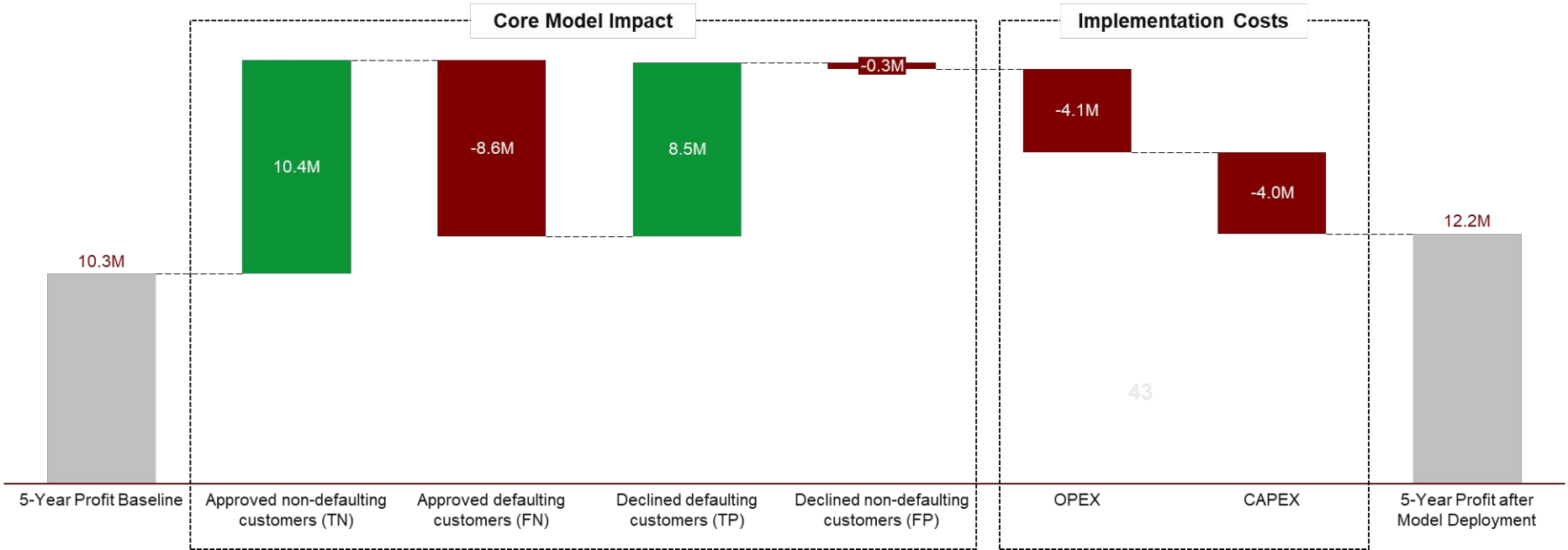


Insights

- True Positives**
 - *Meaning:* **Declined defaulting** customers
 - *Financial Impact:* Cost savings from avoiding paying acquisition costs to defaulting customers
- False Positives**
 - *Meaning:* **Declined non-defaulting** customers
 - *Financial Impact:* Lost revenue opportunity
- False Negatives**
 - *Meaning:* **Approved defaulting** customers
 - *Financial Impact:* Incurred acquisition costs of churned customers
- True Negatives:**
 - *Meaning:* **Approved non-defaulting** customers
 - *Financial Impact:* Identified profit-generating customers

The Random Forest model is expected to create a \$2 million opportunity by increasing profits by 20% for new customers

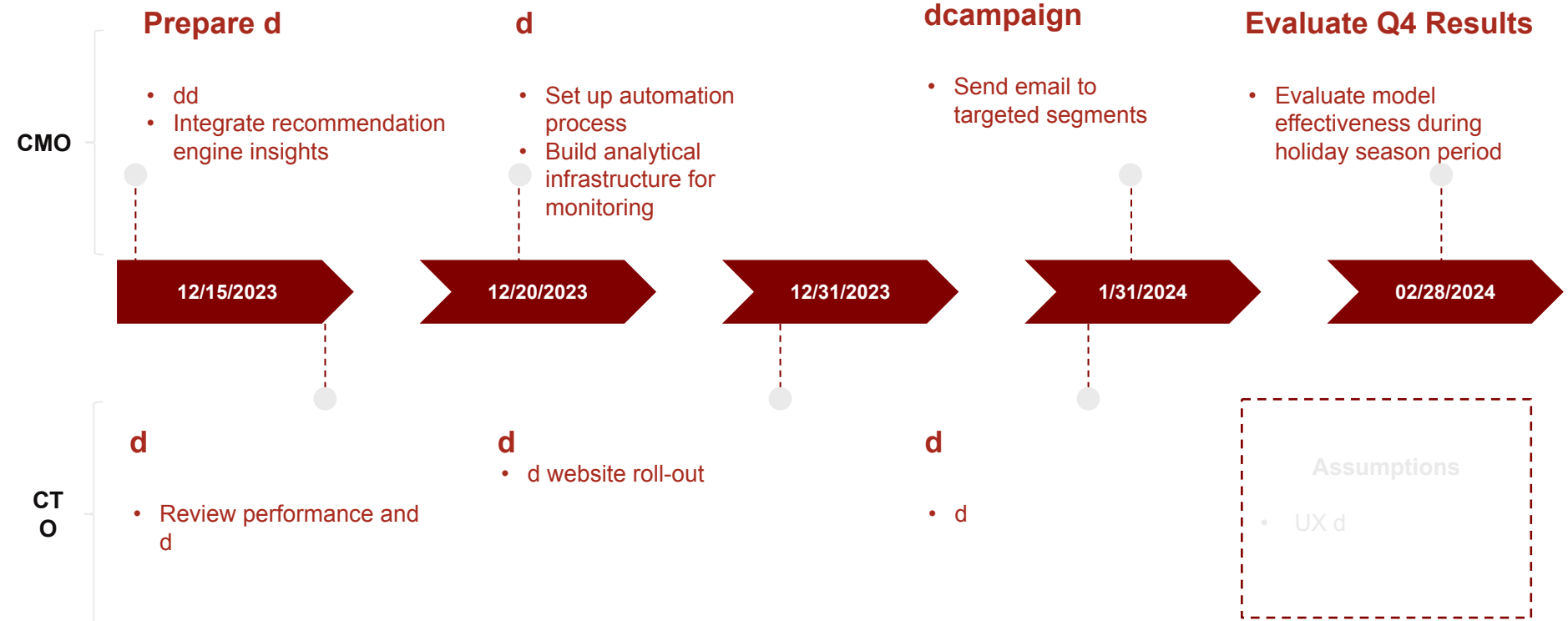
Random Forest Financial Impact
5-year forecast, values in USD



Benefits and Costs Relevant to Model Implementation

| | Tangible | Intangible |
|----------|--|---|
| Benefits | <ul style="list-style-type: none">• Revenue increase• Increased market share• Higher customer lifetime value | <ul style="list-style-type: none">• Operational efficiencies• Increased competitive advantage• Improved decision making |
| Costs | <ul style="list-style-type: none">• Software and hardware cost• Training/implementation cost | <ul style="list-style-type: none">• Employee resistance• Regulatory and compliance risks – ethical use of AI |

Implementation: dddd Timeline through December 2023



ddd

1

ddd

dd

2

ddd

ddd

3

dd

ddd

4

ddd

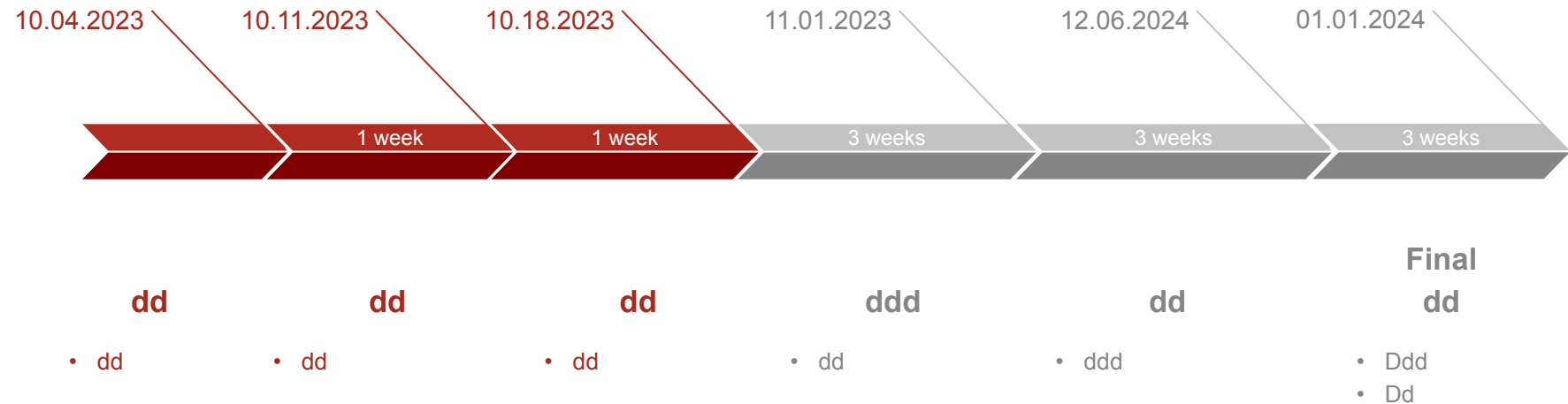
CMO &
Webmaddddd
ster

dd

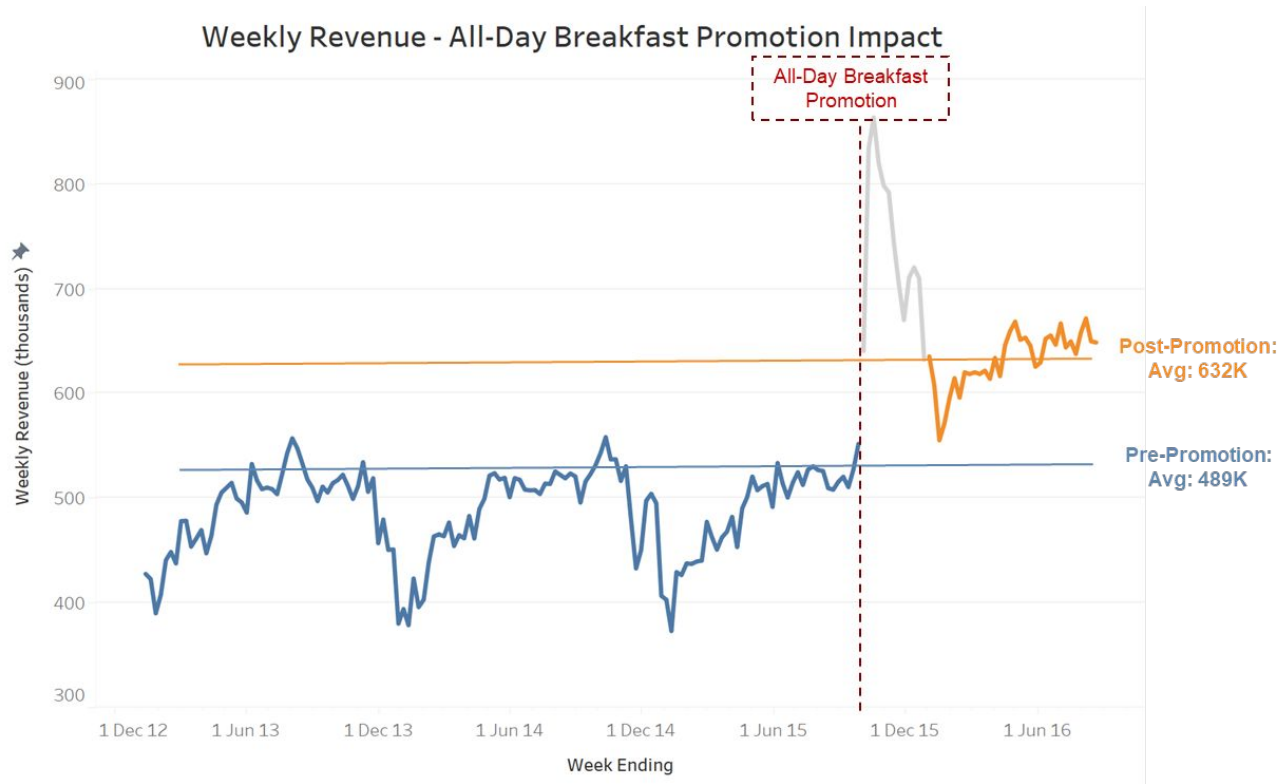
dd

dddd

Preliminary Plan



dddd



All-Day Breakfast Promotion:

↗ ddd

↗ dd

dd

- I. **dd**
 - dd
- II. **dd**
 - dd
- III. **dd**
 - d
 - Dd
 - Dd