

Let's
talk about
what's
possible.[™]

Project Week Presentation

January 19th, 2024

Team 17: Data Lucifer
Louis Lu, Vanessa Tao, Derek Quan, Jinglin Xu



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Let's
talk about
what's
possible.[™]

CONTENTS



Exploratory
Data Analysis



Data
Preprocessing



Data
Modeling



Model Training
& Evaluation



Business
Analysis



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Let's
talk about
what's
possible.[™]

Exploratory Data Analysis

Words & Labels



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

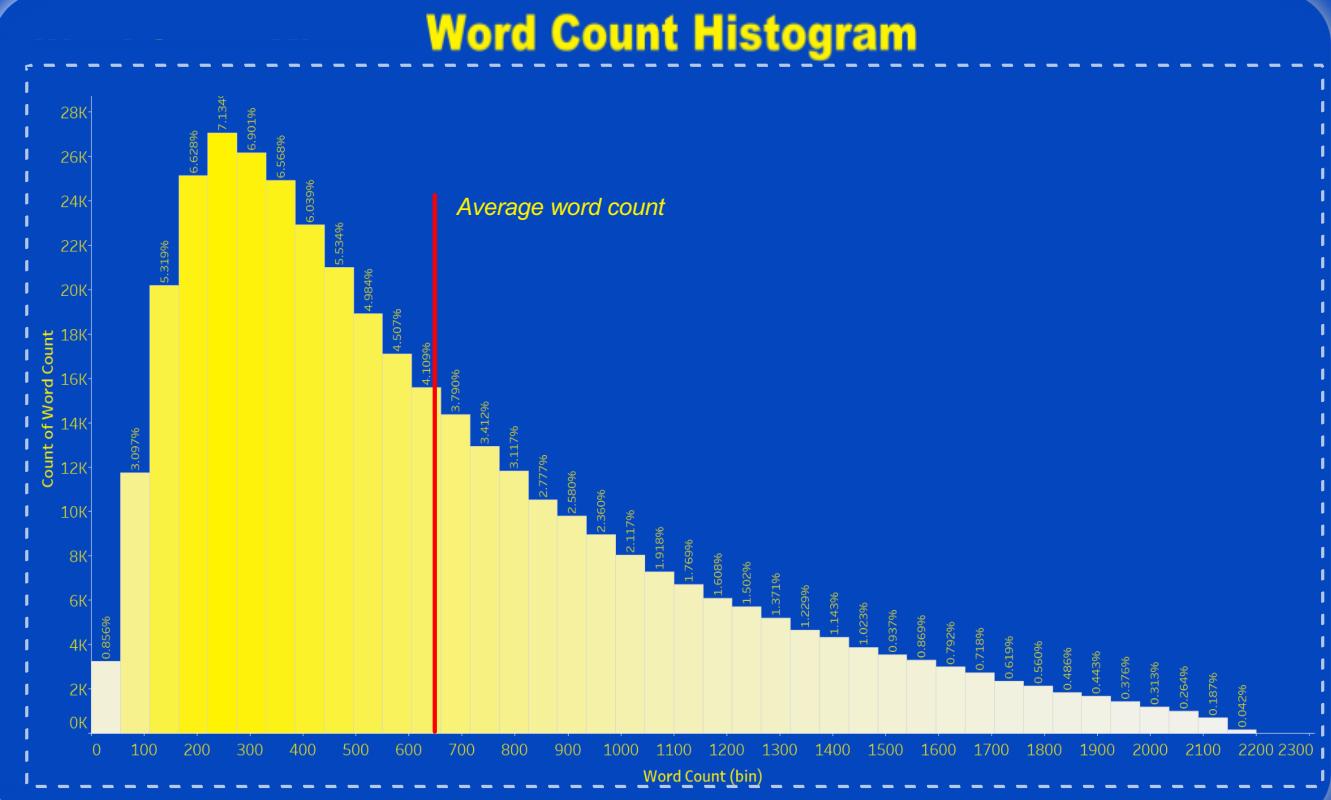
Exploratory Data Analysis: Words

01

Issue of
extensive text

02

Necessity
of text
preprocessing



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

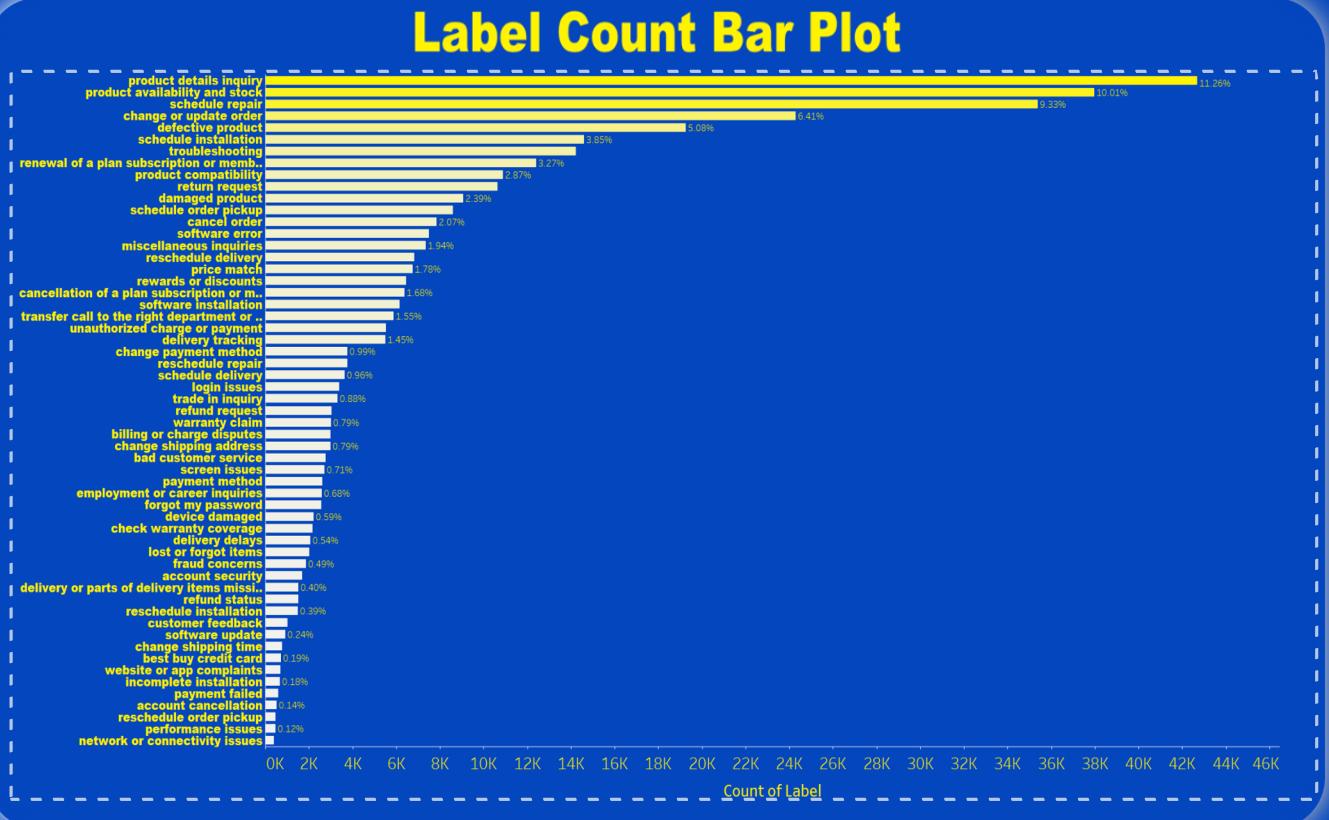
Exploratory Data Analysis: Labels

01

Sign of imbalanced data

02

Suggestion of using F1 score for model comparison



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Let's
talk about
what's
possible.[™]

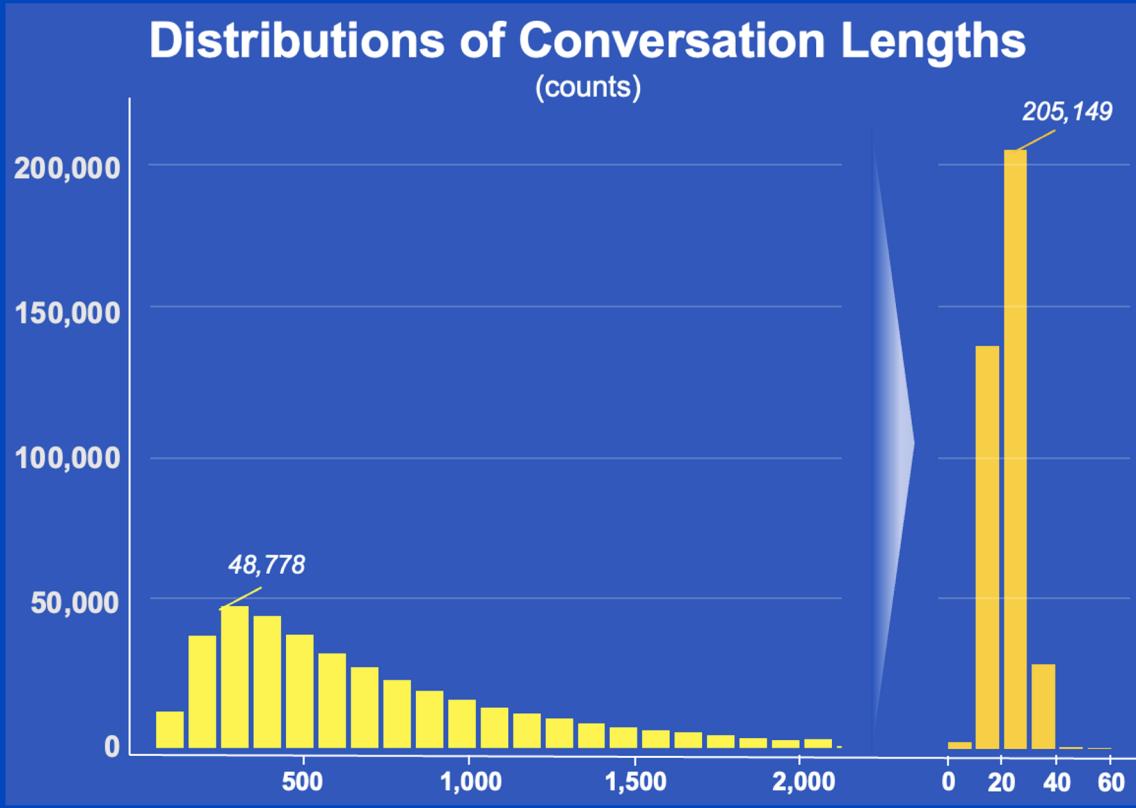
Data Preprocessing

Seq2seq text summarization



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Data Preprocessing: Facebook.bart.xsum



After Summarization

Average Length

Conversation words have been reduced from **638.88** to **22.58**

Text Format

Typos, improper punctuation, and wrong structure have been solved

Information Retention

Redundant sentences and words have been removed to ensure clarity and conciseness



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Data Preprocessing: Facebook.bart.xsum

Raw Data redundant and useless information and typos, adding to modeling difficulty

"agent says for contacting best buy this is name before we get started may i please have your phone number and full name in case or call the disconnected..customer says sure xxx.customer says xxxxxxxx.agent says i appreciate it and may i please also have your name..customer says name..agent says thank you so much name for that important information let me know how can i help you..customer says um i recently just purchased a time for the uh ring doorbell i had it installed date from my parents..agent says okay..customer says they needed the actual time i just plugged it into the wall we paired it. its ringing on my phone but its not ringing in the house when i hit the doorbell.customer says so how do we what how do we get it the test worked we the the chime and the the plugged in chime does work but.customer says it it. it worked for the test when we did the test but it doesn't work when i hit the doorbell.agent says mm okay thank you. okay thank you so much for letting us know well. i need to connect you over to one of our technician..customer says okay great. thank you..customer says yep..agent says um and they will be able to guide you to make it ring okay. please stay on the line. i appreciate it..customer says you think if you."

Summary clean, concise, efficient for modeling

"A customer called best buy had a problem with their doorbell, so they sent a technician to help her out."



After Summarization

Average Length

Conversation words have been reduced from **638.88** to **22.58**

Text Format

Typos, improper punctuation, and wrong structure have been solved

Information Retention

Redundant sentences and words have been removed to ensure clarity and conciseness

Data Preprocessing: Facebook.bart.xsum

Raw Data redundant and useless information and typos, adding to modeling difficulty

"agent says for contacting best buy this is name before we get started may i please have your phone number and full name in case or call the disconnected..customer says sure xxx.customer says xxxxxxxx.agent says i appreciate it and may i please also have your name..customer says name..agent says thank you so much name for that important information let me know how can i help you..customer says um i recently just purchased a time for the uh ring doorbell i had it installed date from my parents..agent says okay..customer says they needed the actual time i just plugged it into the wall we paired it. its ringing on my phone but its not ringing in the house when i hit the doorbell.customer says so how do we what how do we get it the test worked we the the chime and the the plugged in chime does work but.customer says it it. it worked for the test when we did the test but it doesn't work when i hit the doorbell.agent says mm okay thank you. okay thank you so much for letting us know well. i need to connect you over to one of our technician..customer says okay great. thank you..customer says yep..agent says um and they will be able to guide you to make it ring okay. please stay on the line. i appreciate it..customer says you think if you."

Summary clean, concise, efficient for modeling

"A customer called best buy had a problem with their doorbell, so they sent a technician to help her out."



After Summarization

Average Length

Conversation words have been reduced from **638.88** to **22.58**

Text Format

Typos, improper punctuation, and wrong structure have been solved

Information Retention

Redundant sentences and words have been removed to ensure clarity and conciseness

Data Preprocessing: Facebook.bart.xsum

Raw Data redundant and useless information and typos, adding to modeling difficulty

"agent says for contacting best buy this is name before we get started may i please have your phone number and full name in case or call the disconnected..customer says sure xxx.customer says xxxxxxx.agent says i appreciate it and may i please also have your name..customer says name..agent says thank you so much name for that important information let me know how can i help you..customer says um i recently just purchased a time for the uh ring doorbell i had it installed date from my parents..agent says okay..customer says they needed the actual time i just plugged it into the wall we paired it. its ringing on my phone **but its not ringing in the house when i hit the doorbell.**customer says so how do we what how do we get it the test worked we the the chime and the the plugged in chime does work but.customer says it it. it worked for the test when we did the test but it doesn't work when i hit the doorbell.agent says mm okay thank you. okay thank you so much for letting us know well, **i need to connect you over to one of our technician..**customer says okay great. thank you..customer says yep..agent says um and **they will be able to guide you to make it ring okay.** please stay on the line. i appreciate it..customer says you think if you."

Summary clean, concise, efficient for modeling

"A customer called best buy had a problem with their **doorbell**, so they **sent a technician to help her out.**"



After Summarization

Average Length

Conversation words have been reduced from **638.88** to **22.58**

Text Format

Typos, improper punctuation, and wrong structure have been solved

Information Retention

Redundant sentences and words have been removed to ensure clarity and conciseness

Data Preprocessing: Facebook.bart.xsum

Single Word

the i digitstranscribedby customer uh of agent just says
as email was, that name are um not numberwhom phone calling me so there would for and
not correct herewhich voice a well get yourself addressedly because cause you work anything time
or off about backcouple help themselfs been throughin deep Wittenbergs
but curried but curried but curried but curried

Two-Word Phrases

Three-Word Phrases

100,292 20,752,833

After Summarization

Average Length

Conversation words have been reduced from **638.88** to **22.58**

Text Format

Typos, improper punctuation, and wrong structure have been solved

Information Retention

Redundant sentences and words have been removed to ensure clarity and conciseness

Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

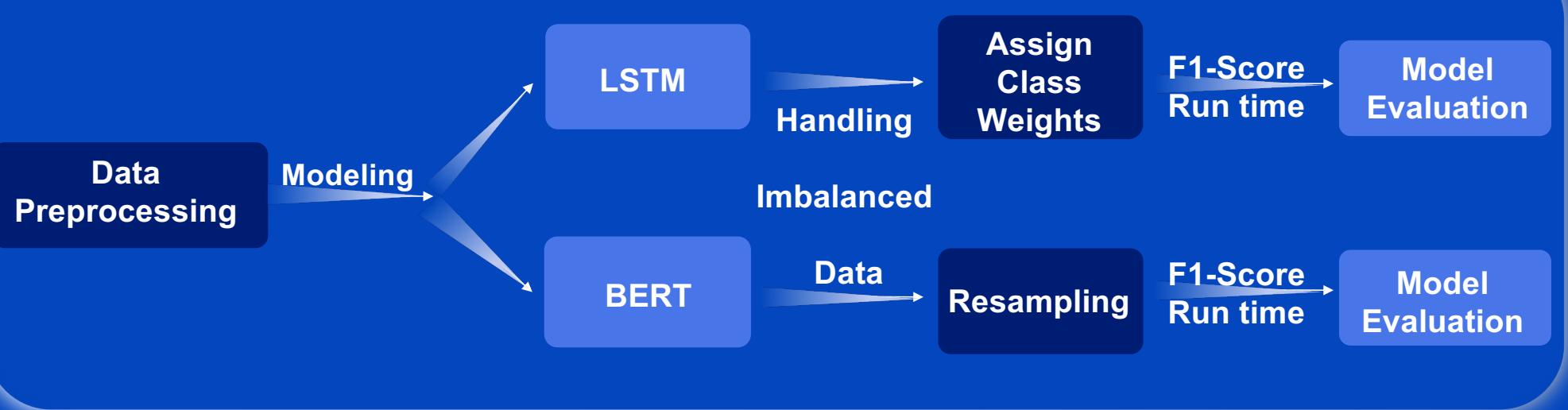
Let's
talk about
what's
possible.[™]

Data Modeling: Tokenization, Imbalanced Data, Albert



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Data Modeling: Tokenization



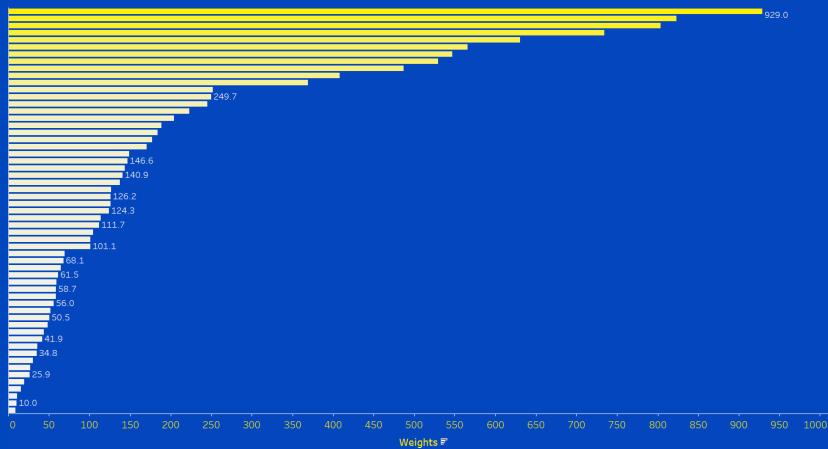
Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Modeling: Imbalanced Data

1

Assign Class Weights

Label Weights Bar Plot



$$\omega_i = \frac{\sum_{m=1}^{57} N_m}{N_i}$$

ω_i : the weight of class i

N_i : the number of text in class i

2

Resampling Labels

- Assign class weights does not work well for Albert
- Balance Class Distribution
- Better Learning of Minority Class
- Simplicity and computational efficiency

Set threshold to 5,000

Set threshold to 10,000

Set threshold to 15,000

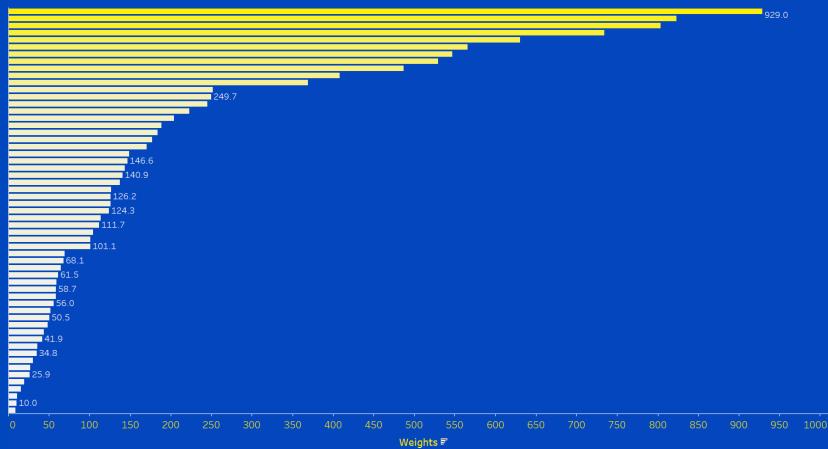


Modeling: Imbalanced Data

1

Assign Class Weights

Label Weights Bar Plot



$$\omega_i = \frac{\sum_{m=1}^{57} N_m}{N_i}$$

ω_i : the weight of class i

N_i : the number of text in class i

2

Resampling Labels

- Assign class weights does not work well for Albert
- Balance Class Distribution
- Better Learning of Minority Class
- Simplicity and computational efficiency

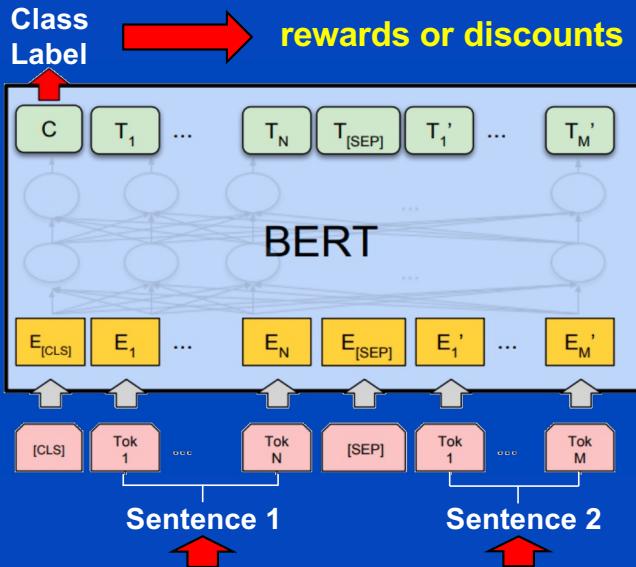
Set threshold to 5,000

478,985 text



Modeling: From Bert to Albert

Bert-based-uncased model



A woman has called best buy to ask if she can get a refund for a watch she bought on date and the cashier rang her up for the full lawn like my best buy total rewards membership thing.

RNN

usually sees the words one after the other

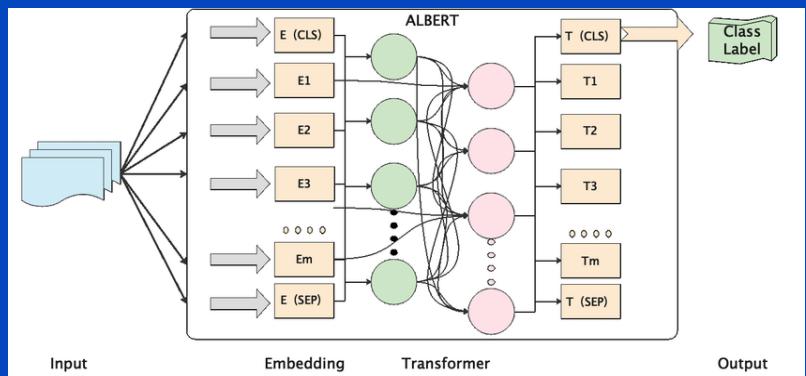
GPT

internally masks the future tokens

BERT

allows the model to learn a **bidirectional representation** of the sentence

A-little-bert model



Model	Parameters	Layers	Hidden	Embedding
BERT	108M	12	768	768
ALBERT	12M	parameter sharing	12	768

- ✓ **Solve** the issue of training time exponentially increasing with the number of epochs
- ✓ **Improve** the validation performance



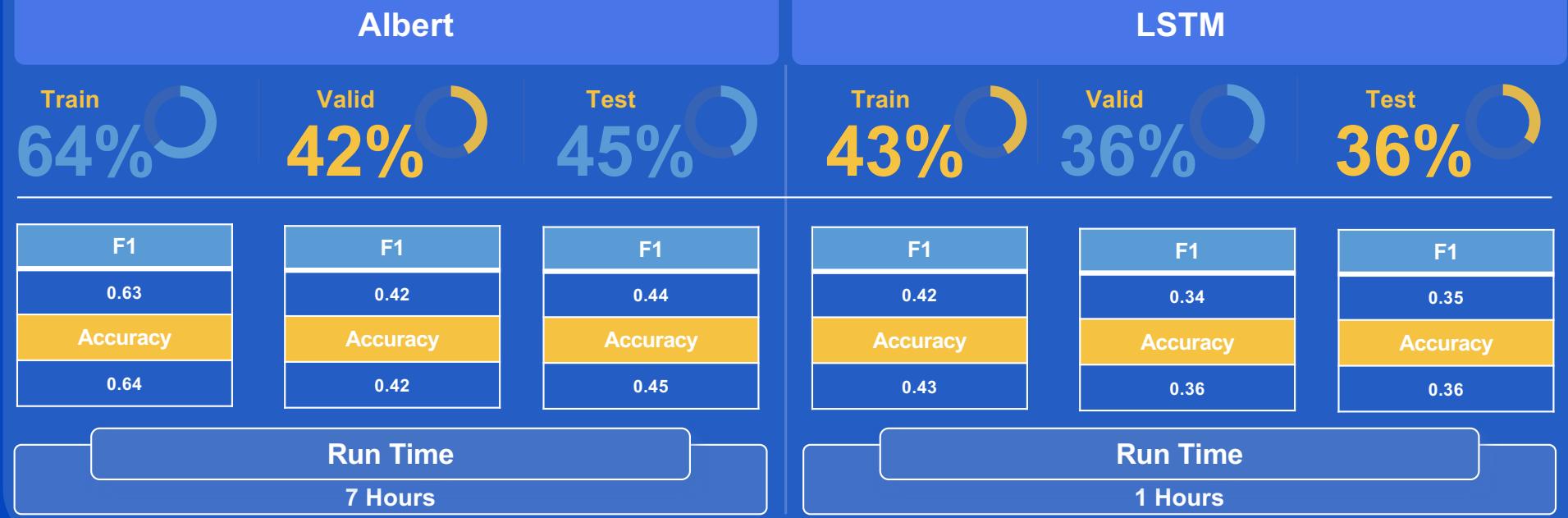
Let's
talk about
what's
possible.[™]

Model Training & Evaluation: Albert & LSTM



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Model Training & Evaluation:



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

Let's
talk about
what's
possible.[™]

Business Analysis

SWOT Analysis & Business Strategy



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

SWOT Analysis

- Enhance BestBuy practical experience in labeling predictions
- Assist BestBuy in resource allocation strategies and proactive problem-solving to stay ahead of challenges

- Highly imbalanced data, hard for model to train
- Over-Reliance on Automation

Strengths

Weaknesses

Opportunities

Threats

SW
OT

- Provide better personalized customer service
- Strategic partnership, enhancing bargaining power

- Competitors
- Data security



Business Strategy: Resource Allocation

“One-size-fits-all” model



State	Locations	Percentage	Population Per Location
California	143	13.8%	276,491
Texas	106	10.2%	274,957
Florida	65	6.3%	331,356
New York	53	5.1%	381,429
Illinois	44	4.2%	291,193

Top 5 states with most Best Buy stores



Data source: BEST BUY. Method, coding, and presentation: Data Lucifer Team.

“Geo-Targeted Retail ” model

1

Add location data to phone records enables us to identify problem areas

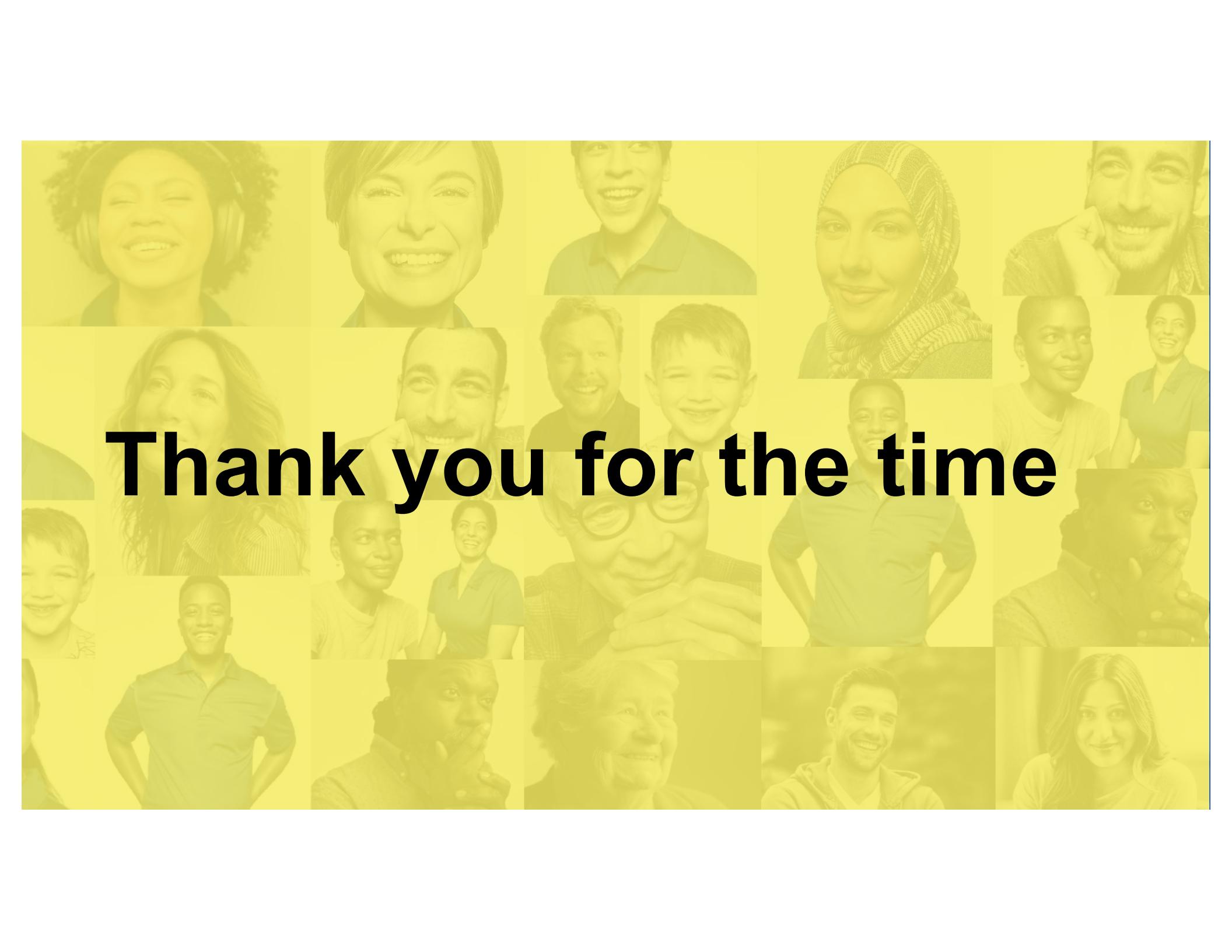
2

Tailor improvements to branches and logistics

3

Optimization based on location-specific issues

Store Efficiency & Customer Satisfaction



Thank you for the time