

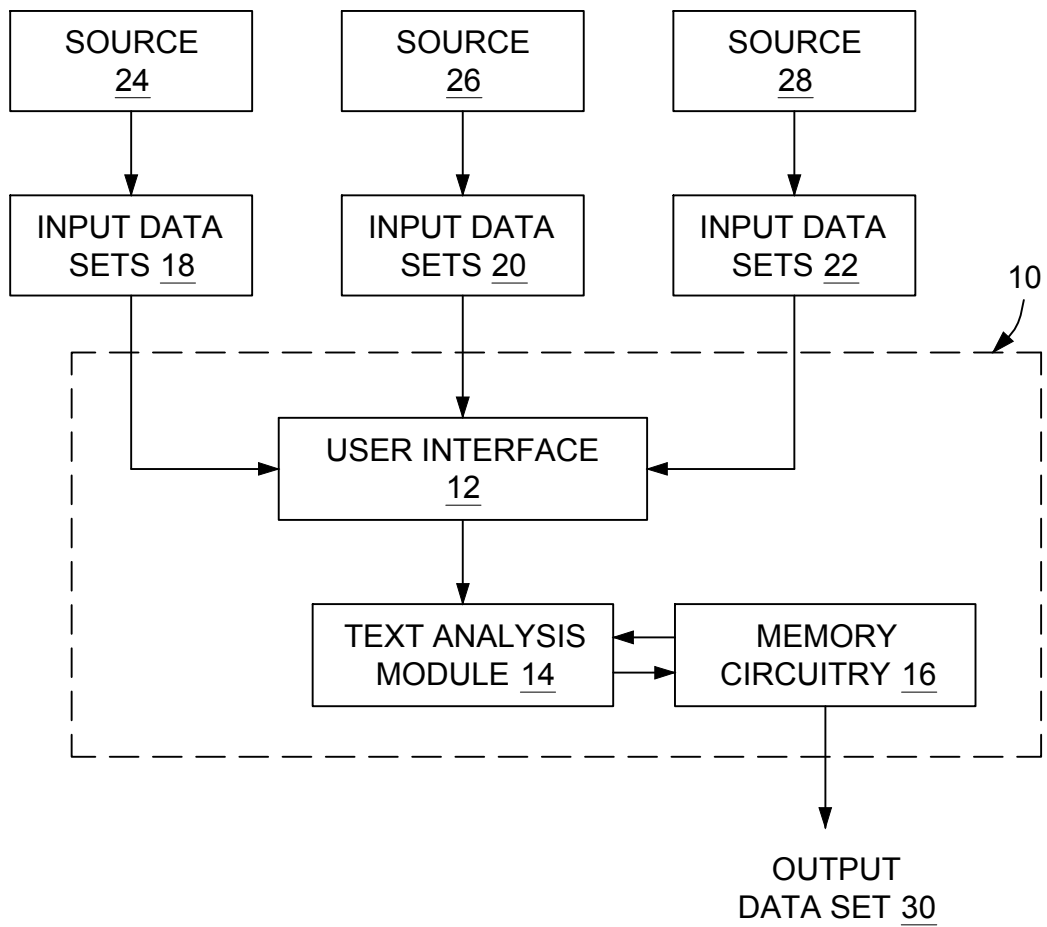


FIG. 1

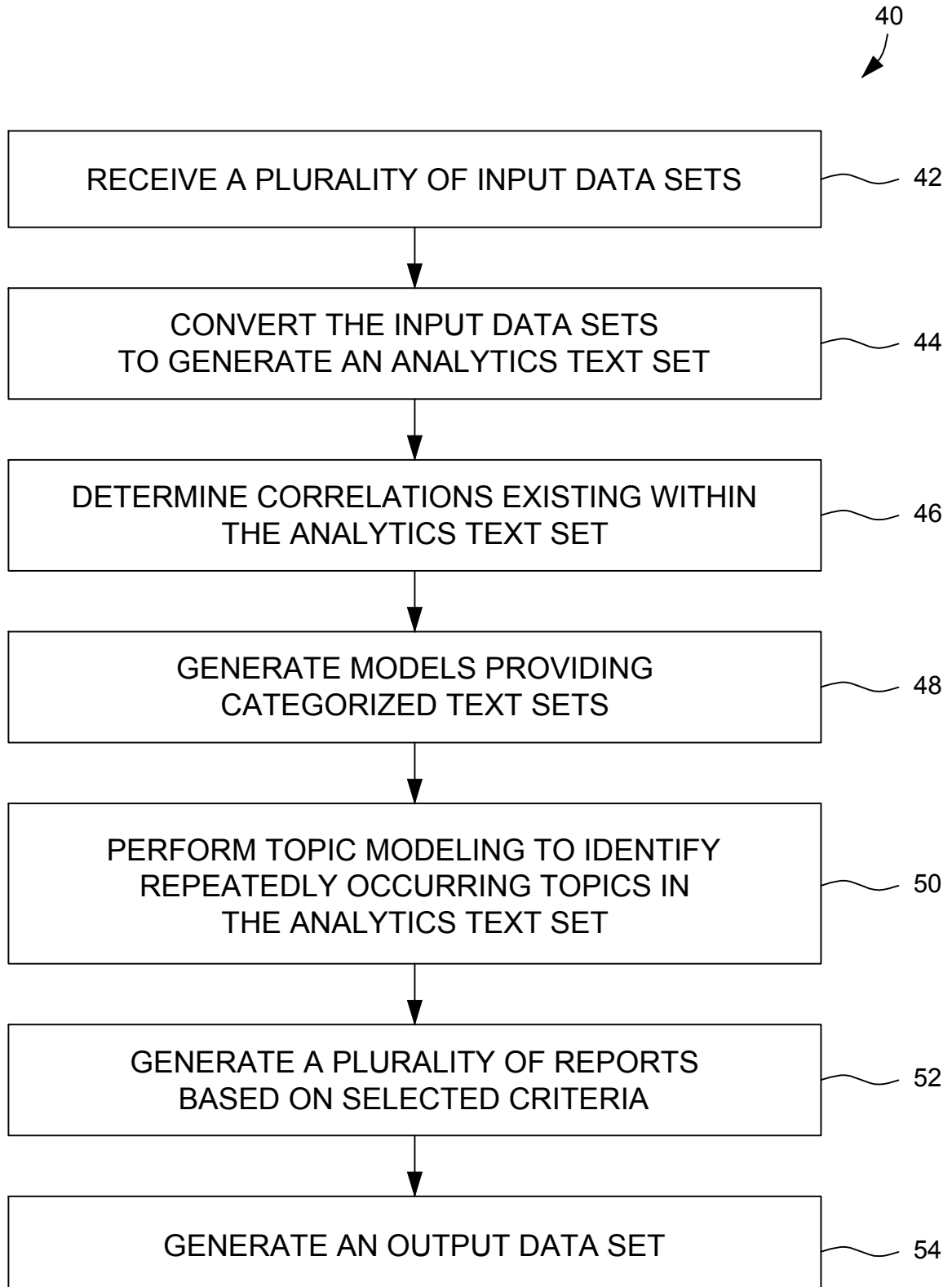


FIG. 2

60 ↗

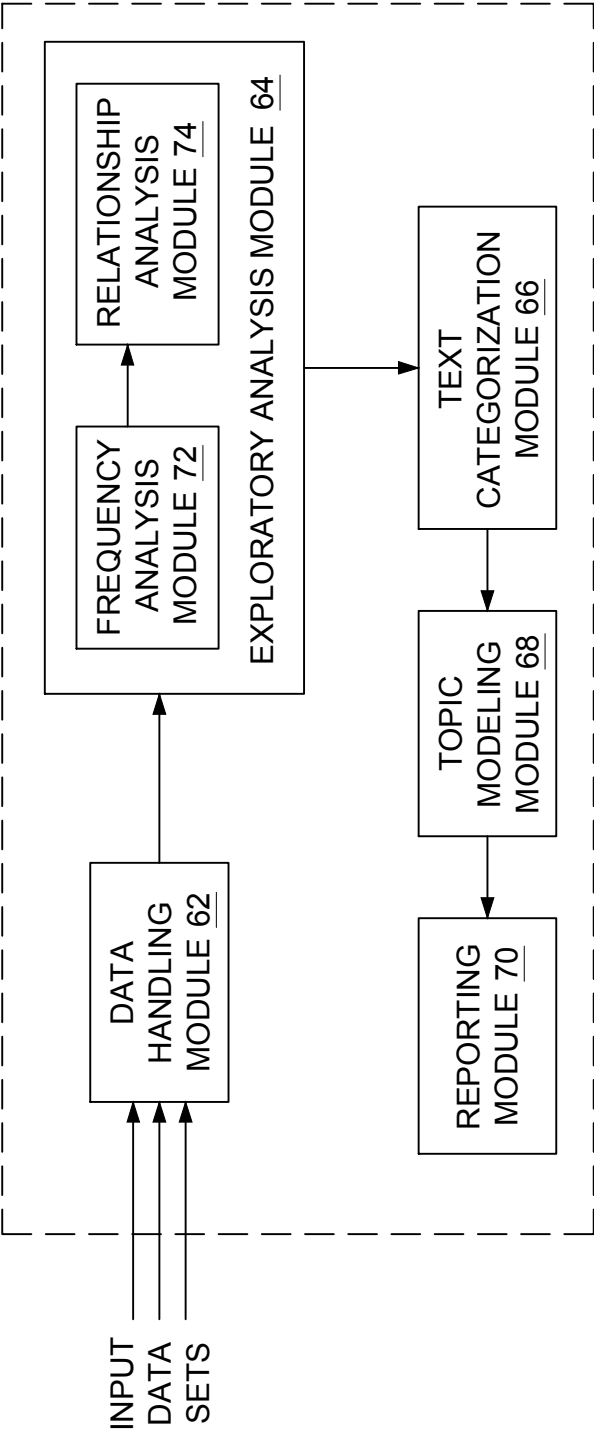


FIG. 3

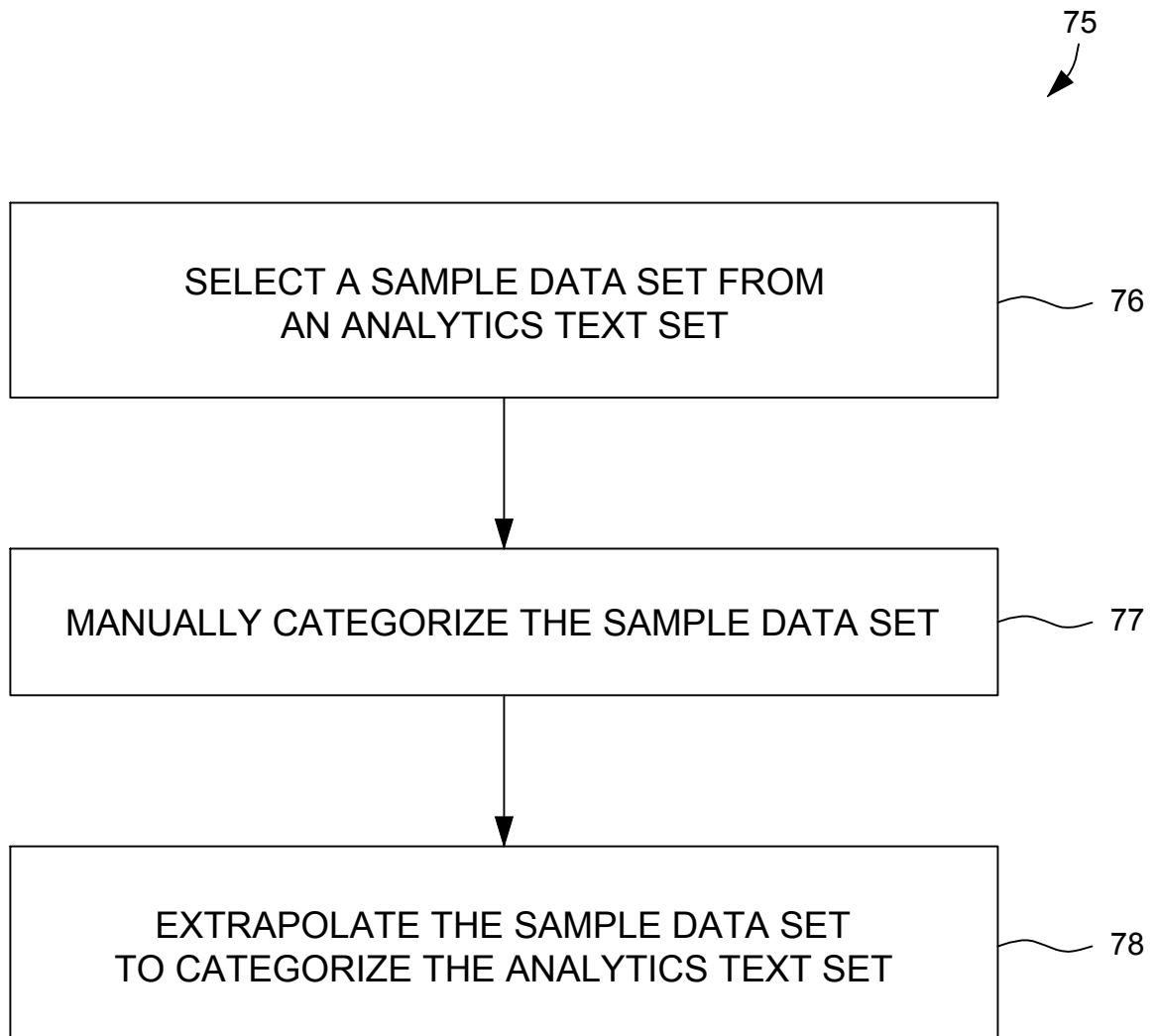


FIG. 4

80

82

Project Setup

Data Preparation

Data Quality Analysis

EDA

Modeling

Segmentation

Reporting Framework

Text-Mining

Add Dataset

Variable Categorization

Dataset View

Data Dictionary

Business Categorization

Scenario Builder

Add Dataset

Import from Data Source

Add Dataset

Dataset Path:

Browse

Note: If using Windows 7 or higher, datasets places on the desktop cannot be accessed

Add Dataset

Import Dataset

Datasets

☐ Name	Observations	Variables	Size (MB)	Compressed
☐ competitors_of_eliquis_tm	273	2	0.03	No
☐ perception_data_tm	1028	2	0.13	No
☐ prescription_decisions_tm	904	2	0.11	No
☐ sources_tm	228	2	0.02	No
☐ type_of_patients_tm	595	2	0.06	No
☐ unmet_needs_tm	71	2	0	No

84

86

FIG. 5

90

92

98

94

96

97

100

102

104

106

108

Project Setup		Data Preparation	Data Quality Analysis	EDA	Modeling	Segmentation	Reporting Framework	Text-Mining
Data Handling		EDA Text	Text Categorization	Topic Modeling				
Data Pre-Processing		Observation split						
Report Generation		Report Viewing						

Dataset: shorter dataset

☐ Panel Levels

☒ Across Datasets

Overall_Rating sent_Overall_Rating

Levels

☐ 0|negative

☐ 2|neutral

☐ 3|positive

☐ 4|positive

Variable Panel

☐ Variable

☒ Comments

☐ Overall_Rating

☐ sent_Overall_Rating

Categorical Variables

View Data

Analysis Language: English

Select Language

English

French

German

Portuguese

Spanish

Dataset View

Search

Comments

I just spent at least 25 morning at the store. That is unacceptable. Check that stores tapes. I'm in a grey jacket and a black hat. The guy in front of me had a produce item the checker couldn't find on his sheet. After spending a very long time looking at the sheet trying different numbers and never doing a price check or asking for help he finally gets the person in front of me checked out only to realize that he messed up. D

at 2:00am in the

It's horrible long waits - they have 29 cashier spots open and on Friday night (8 pm) and Sat/Sunday only have 8 cashiers open - and I waited 35 minutes in line to buy 9 or 10 items and no self checks(he told me it's the 2nd highest theft store in area so not self checks) every time I come in here I must wait 22-35 minutes to buy something - I can't waste that time - I have children and family- i

Reports

Report Name

Report Detail

Delete

Page 1

100

Rows/Page

Create indicator

GO

FIG. 6A

110

Project Setup

Data Preparation

Data Quality Analysis

EDA

Modeling

Segmentation

Reporting Framework

Text-Mining

Data Handling

EDA Text

Text Categorization

Topic Modeling

Data Pre-Processing

Observation split

Report Generation

Report Viewing

Dataset: shorter dataset

Analysis Language: English

☐ Panel Levels

☒ Across Datasets

Overall_Rating sent_Overall_Rating

Levels

☐ 0|negative

☐ 2|neutral

☐ 3|positive

☐ 4|positive

Variable Panel

☐ Variable

☒ Comments

☐ Overall_Rating

☐ sent_Overall_Rating

Dataset View

Data Cleaning

Data Extraction

Data Cleaning Operations

Stem Words

☐ Porter's Stem

Add Operation

Report Name:

Operations

Submit

Reports

Report Name

Report Detail

Delete

View Data

Categorical Variables

FIG. 6B

120

122

124

Project Setup		Data Preparation		Data Quality Analysis		EDA		Modeling		Segmentation		Reporting Framework		Text-Mining	
Data Handling		EDA Text		Text Categorization		Topic Modeling									
Data Pre-Processing		Observation split													
Report Generation		Report Viewing													

Dataset

shorter dataset

Treatment Process

Across Dataset

126

Variable Panel

☐	Variable	Label
☒	Comments	132
☐	Overall_Rating	134
☐	sent_Overall_Rating	136

☒ Categorical Variables

Split Options

128

Split on Variable

Comments

130

Delimiter

Select Delimiters

☒ Minimum length to split

20

Minimum length after split

3

Preview

0

Split Preview

138

Comments

Split_Comments

Split Variable Name

New Dataset Name

Split Observations

Reports

Report Name	Report Detail	Delete

FIG. 6C

Figure 1 is a screenshot of a software interface for data analysis. The interface is divided into several sections. At the top, there are tabs for "Data Handling", "EDA Text", "Text Categorization", and "Topic Modeling". Below these, there are sections for "Frequency Analysis" (152), "Relationship Analysis" (154), and "Report Generation" (156). The "Report Generation" section includes a "Report Viewing" button (158). The main area displays a "Dataset" (160) named "surface-final clean - sample_n...". It shows a list of variables (162) including "Availability", "Promotional", "Anno", "Category", "Category_number", "Comparison", "competition", "Intent_to_Purchase", "Posts", "Price", "Reviews", "Specifications", "Accessories", and "Updates". A "Variable Panel" (164) is shown on the right, listing these variables. A "Panel Levels" section (166) shows a list of categories: "Levels", "Comparison/competition", "Reviews", "Specifications/Accessories/Update", "Availability/Promotional/Announce", "Intent to Purchase", and "Price". A "Categorical Variables" section (168) is also present. At the bottom, there is a "Reports" section (170) with a table showing "Report Name" (one), "Report Detail" (posts), and "Delete" (⊖).

FIG. 7

180

Project Setup

Data Handling

Frequency Analysis

Report Generation

Reports

Data Preparation

EDA Text

Relationship Analysis

Report 1

Data Quality Analysis

Text Categorization

Options: Noun:true, Verb:false, Adjective:true | Type of Analysis: Min Frequency:1, Max Frequency:Inf, Top Terms:100 |

EDA

Topic Modeling

Report Viewing

Modeling

Segmentation

Reporting Framework

Text-Mining

Chart

Cloud

Table

182

Selected Variables

dvde_CONTENT

184

Keyword

loan

card

credit

car

bank

job

hous

home

account

time

day

186

Frequency

566

555

523

481

440

373

364

360

300

278

270

190

Frequency Share (in %)

1.53

1.5

1.42

1.3

1.19

1.01

0.98

0.97

0.81

0.75

0.73

192

No. of Comments

523

507

495

472

451

349

359

332

288

260

551

194

Comment Share (in %)

8.83

8.56

8.35

7.96

7.61

5.89

6.06

5.6

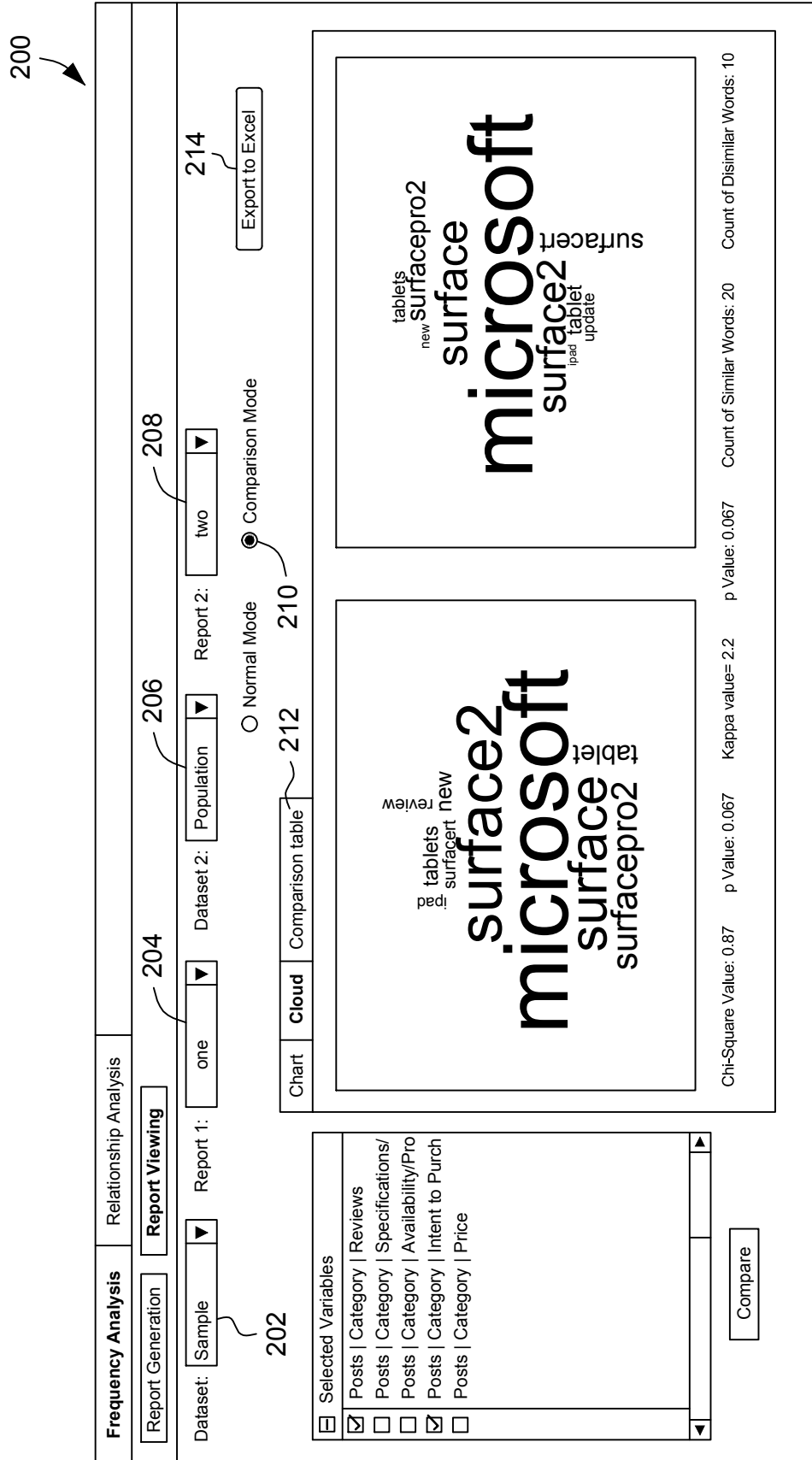
4.86

4.39

9.3

196

FIG. 8A



222

224

226

228

230

220

Project Setup	Data Preparation	Data Quality/Analysis	EDA	Modeling	Segmentation	Reporting Framework	Text-Mining
Data Handling	EDA Text	Text Categorization		Topic Modeling			
Model Definition		Model Diagnostics	Predict	Iteration History			

Dataset

training_dataset

Variable Panel

Label	Variables
☐	ARTICLE_ID
☐	ARTICLE_URL
☐	AUTHOR
☐	BLOG_POST_SENTIMENT
☐	COMMENT_COUNT
☒	CONTENT
☐	ENGAGEMENT
☐	FOLLOWERS
☐	FOLLOWING
☐	FORUM_THREAD_SIZE
☐	HEADLINE

Select Dependent Variables

Selection	Variable	Remove
☒	Lead.Identification	☐

Options

Algorithms

☒ SVM
☒ Random Forest
☒ GLMNET
☒ Maximum Entropy

Select All
Deselect All

Property

Minimum Word Length 3
Minimum Doc Frequency 1
Weighting TF
☒ Remove Punctuation
☒ Strip Whitespace
☒ Convert to Lower Case
☒ Remove Numbers
☒ Stem words
☒ Remove Stop Words

Select All
Deselect All

Validation Scenario

Build Model

FIG. 9

240

242

Project Setup

Data Handling

Model Definition

Data Preparation

EDA Text

Model Building

Data Quality Analysis

Text Categorization

Model Diagnostics

EDA

Topic Modeling

Predict

Reporting Framework

Segmentation

Modeling

Text-Mining

Iteration History

Iteration

246

244

Dataset

SAMPLE_ONE

Dependent Variable

Category

Iteration

2

Base Iteration: 1

Validation Scenario: Train:Test Ratio :: 80:20

☐	Variables	Category
☐	CiscoPhil:Awesome customer servi	2
☐	negociosdinero: Where can I mail	3
☐	I have an AMex_a B of A and a Cha	2
☐	QUOTE (bemmei jul 24 2010_ 05:49	2
☐	credit001:Chase Credit Card Applic	2
☐	G-spot_ explained the Santa Clara_	4
☐	Secured Loans IStar Hires J.P. Mo	2
☐	So my credit sucks_ I'm working to	3
☐	Yeah Marty's head is still all up in	1

Model Stats

248

Model Statistics

Statistics	Value
Consensus Accuracy	0.575
SVM Accuracy	0.525
Random Forest Accuracy	0.525
Maximum Entropy Accuracy	0.475
GLMNET Accuracy	0.575

Options

Validation Scenario

FIG. 10

250

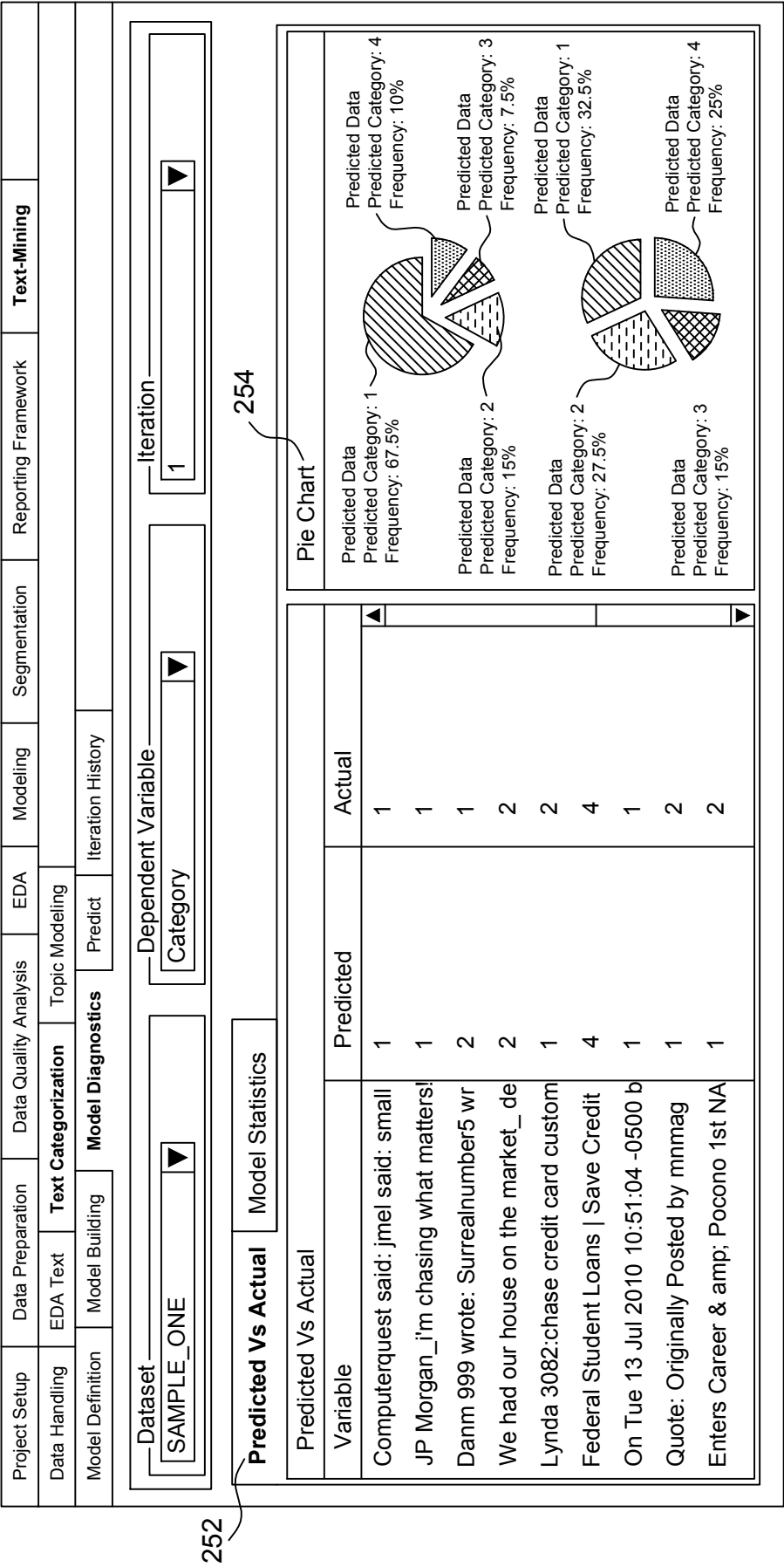


FIG. 11

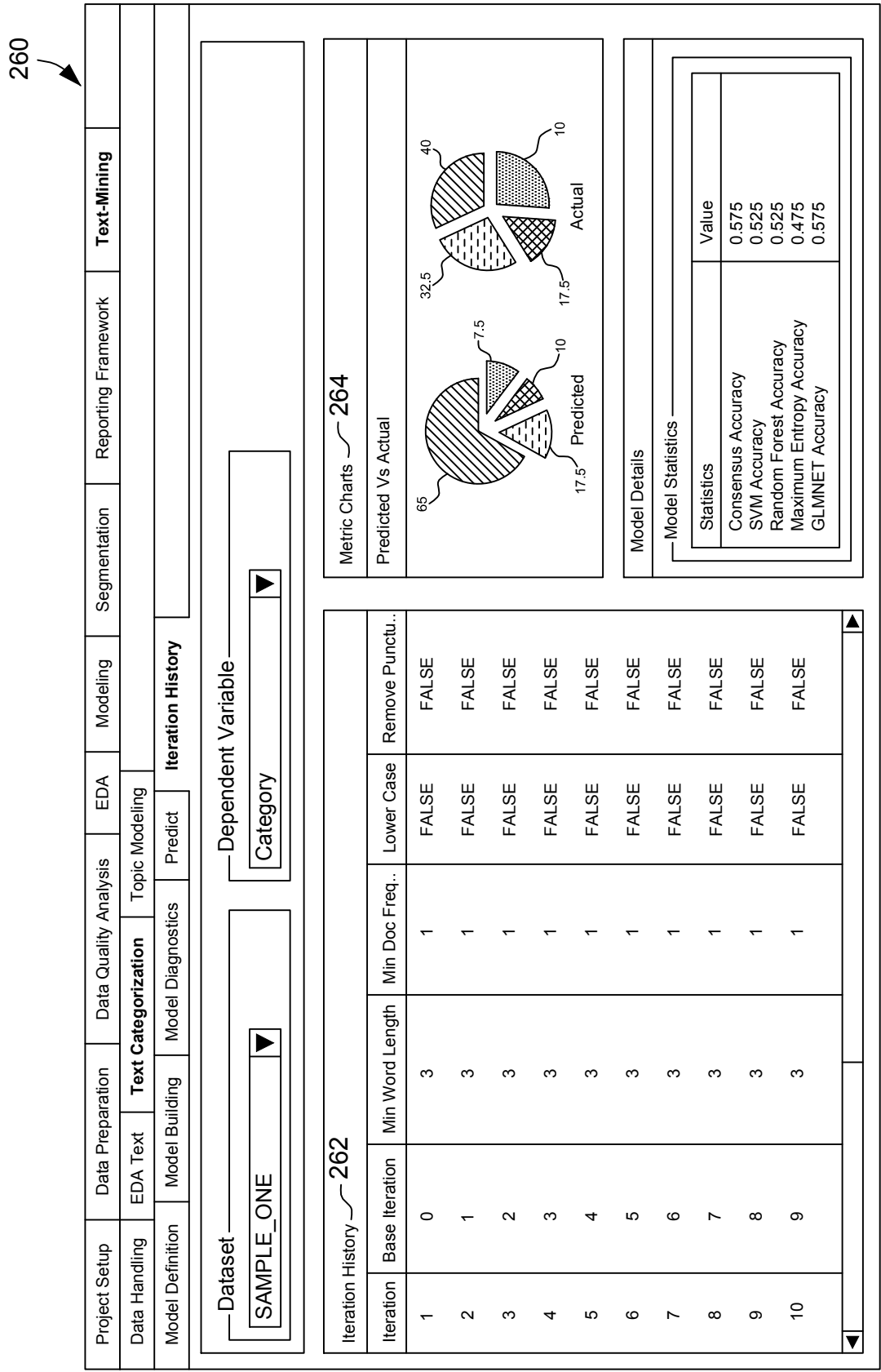


FIG. 12

270

Project Setup

Data Handling

Selection

Data Preparation

EDA Text

Reports

Data Quality Analysis

Text Categorization

EDA

Topic Modeling

Modeling

Segmentation

Reporting Framework

Text-Mining

Dataset

COMPLETE_DATASET

Variable Panel

se.	Variables
☐	Comment
☐	Comment_CleanComment

☒ Parts of Speech

☒ Complete comment

☐ Entity

☒ Input Parameters

☒ Manual

☐ Automatic

No. of Topics:

1

No. of keywords?Topic:

1

No. of Iterations:

1

Submit

Reports

Report Name	Variable Name	Filter	Delete
8topics_4keywords	Comment_CleanC...		☐
5topics_5keywords	Comment_CleanC...		☐
10topicsand_5ley..	Comment_CleanC...		☐
8topics_and_4key..	Comment_CleanC...		☐
5topics_and_5key..	Comment_CleanC...		☐
5keywords_and_5..	Comment_CleanC...		☐
4topics_6keywords	Comment_CleanC...		☐
4Topics_6Kwords	Comment_CleanC...		☐
5KeyWwords_5topic	Comment_CleanC...		☐

FIG. 13

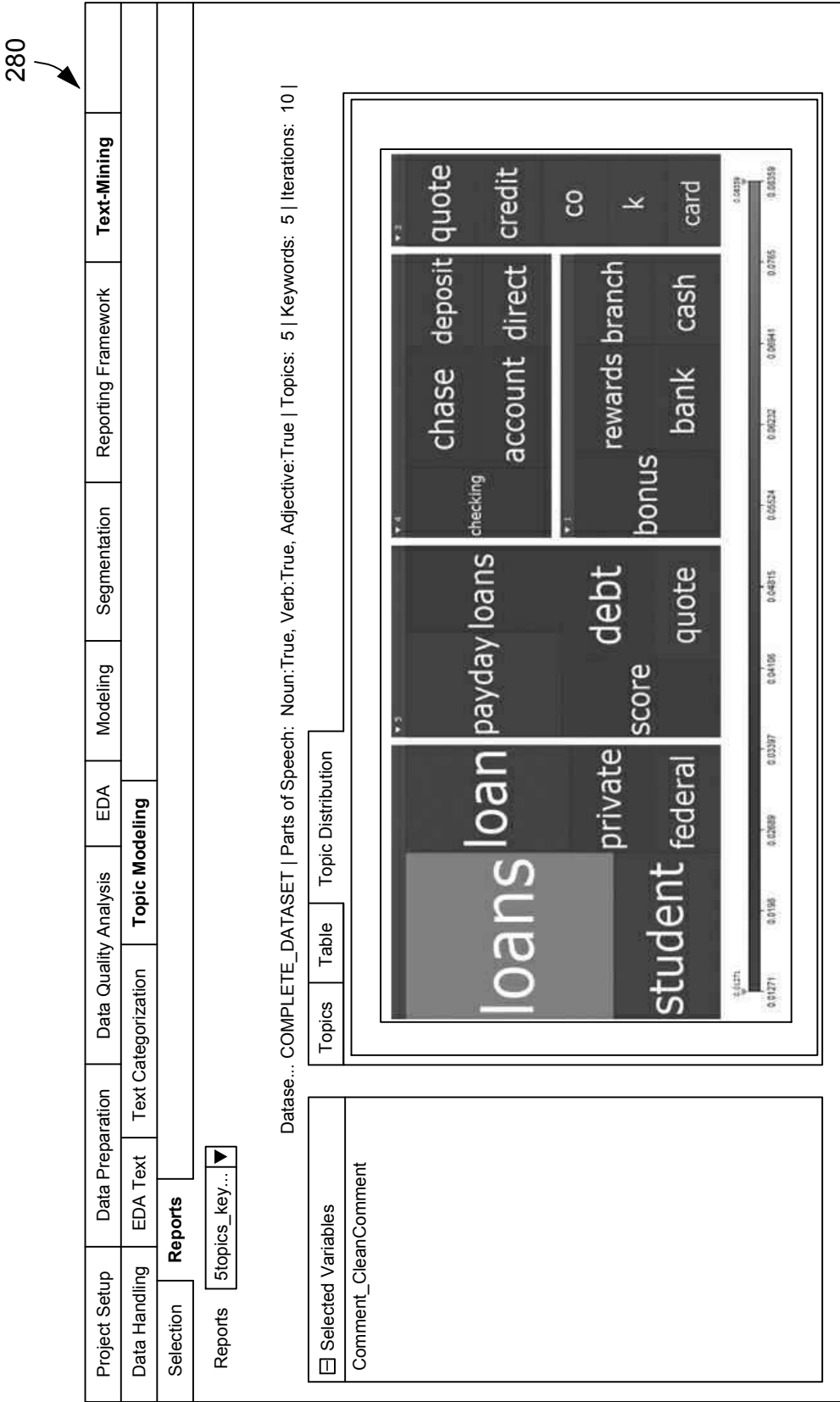


FIG. 14

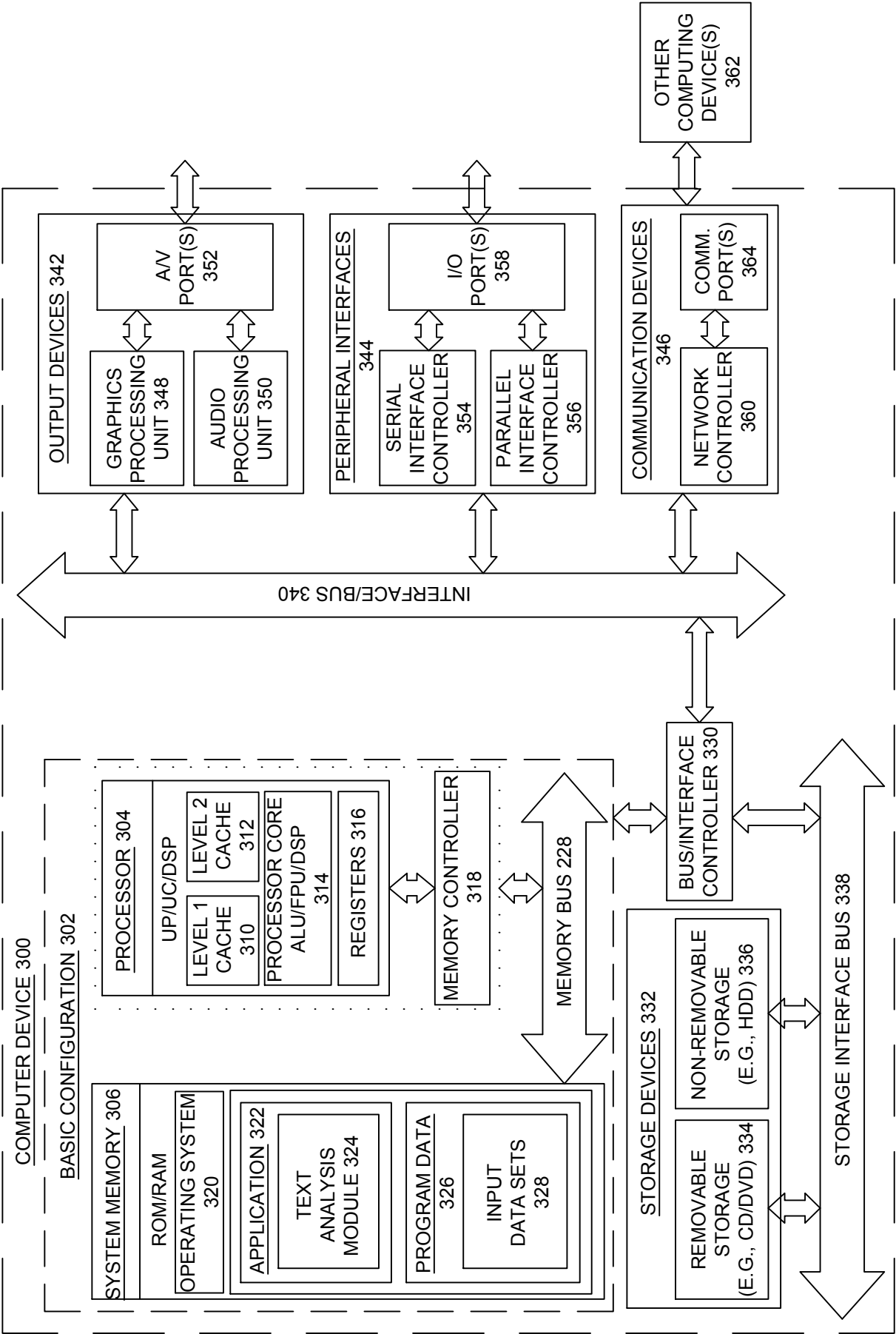


FIG. 15