

Second Language Acquisition Modeling

An analysis of data from Duolingo

By Kate Christensen



What is Duolingo?

Home page
for a course a
user is
enrolled in:

The screenshot shows the Duolingo home page for a German course. The top navigation bar includes links for LEARN, STORIES, DISCUSS, SHOP, and MORE. The user's profile is shown with a German flag, a crown icon, and a score of 449. The main content area displays a grid of lesson icons, each with a yellow circular background and a small yellow tag indicating the number of lessons completed. The lessons are arranged in rows: Basics 2 (5 lessons), Phrases (5 lessons), Idioms (1 lesson), Flirting (1 lesson), Christmas (1 lesson), Acc. Case (5 lessons), Intro (5 lessons), and Food 1 (5 lessons). A blue button labeled 'TRY PLUS' is located next to the Phrases lesson. On the right side, there is a 'Daily Goal' section with a large orange circular progress indicator showing 558 DAYS. To the right of the progress indicator, it says '13/10 xp goal met' and '10 hours left'. Below the progress indicator is a line graph showing the user's progress over the last seven days (Th, F, Sa, Su, M, Tu, W). The graph shows a peak on Thursday (Th) at approximately 28, followed by a drop on Friday (F) to approximately 15, a rise on Saturday (Sa) to approximately 20, a drop on Sunday (Su) to approximately 12, and then a relatively flat trend on Monday (M) at approximately 14, Tuesday (Tu) at approximately 14, and Wednesday (W) at approximately 13.

LEARN STORIES DISCUSS SHOP MORE

German 449 558 2049

TRY PLUS

Basics 2 Phrases

Idioms Flirting Christmas

Acc. Case Intro

Food 1

Daily Goal EDIT GOAL

558 DAYS

13/10 xp goal met

10 hours left

Th F Sa Su M Tu W

About the Data

- Each observation is an individual word a user encounters within an exercise
- The particular dataset analyzed was Spanish speakers learning English
- This dataset had over 2.6 million observations

Data Features

- **User data:**

- Unique id for each user
- Countries user has used the app in

- **Exercise data:**

- Exercise format
- Exercise index
- Response time
 - Response times longer than 60 seconds were cut
- Prompt used
- Session type
- Client type

- **Token data:**

- What the word was
- Part of speech
- Morphological info
 - Plural, definite, possessive, etc.
- Syntactic info
 - Not included because previous work on this data indicated a lack of confidence in its accuracy



10 IN A ROW



Translate this sentence

I do not understand the
Japanese class at all. I am in
trouble.

日本語 の 授業 が 全然
分かり ませ ん こまり ます
つまらなかった
つまらない 困り
こまって

You are correct



CONTINUE

Type what you hear



Je peux être un mauvais garçon

Translation:

I can be a bad boy.



CONTINUE



Translate this sentence



Cat.

Cat

You are correct

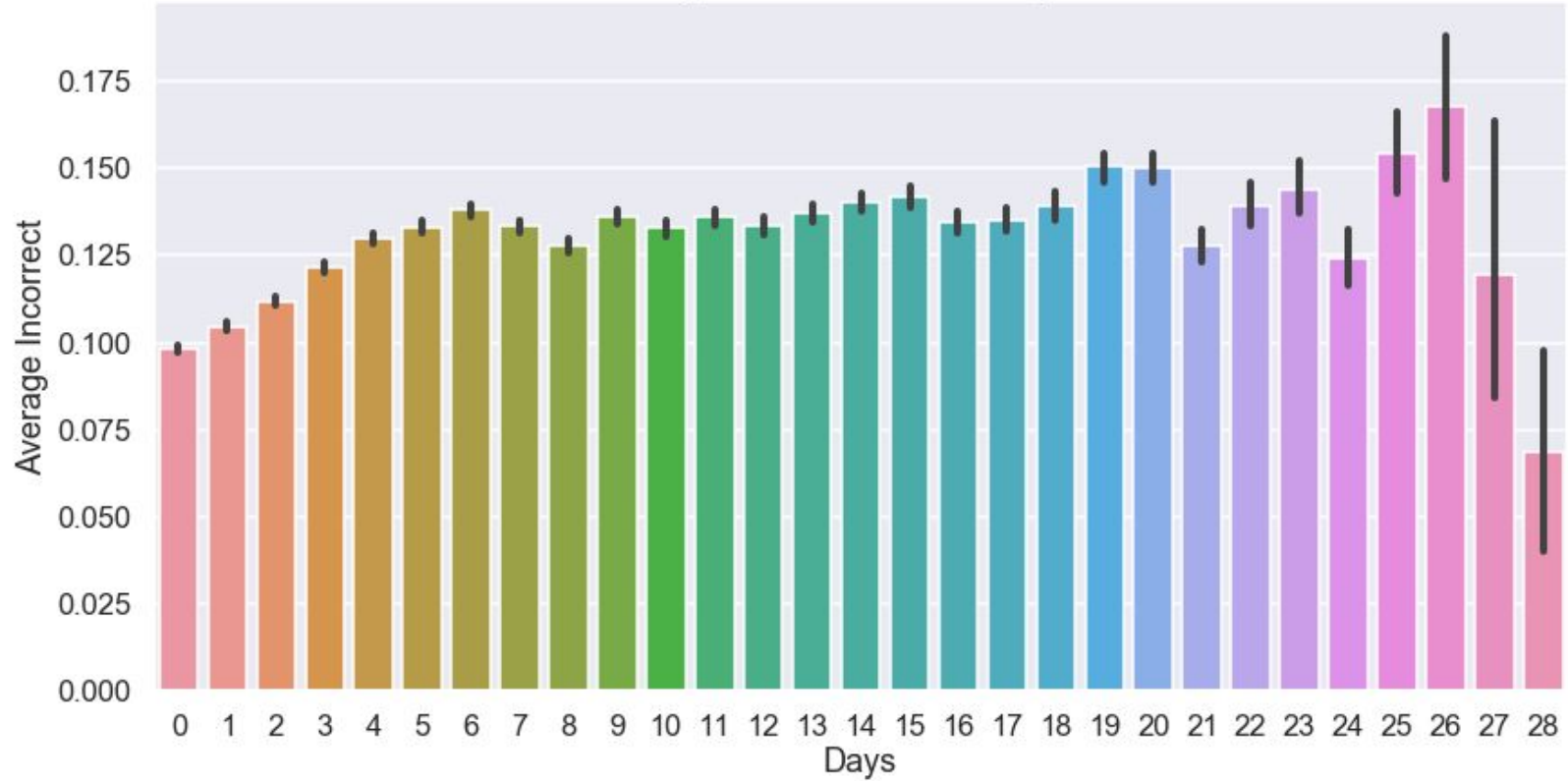


CONTINUE

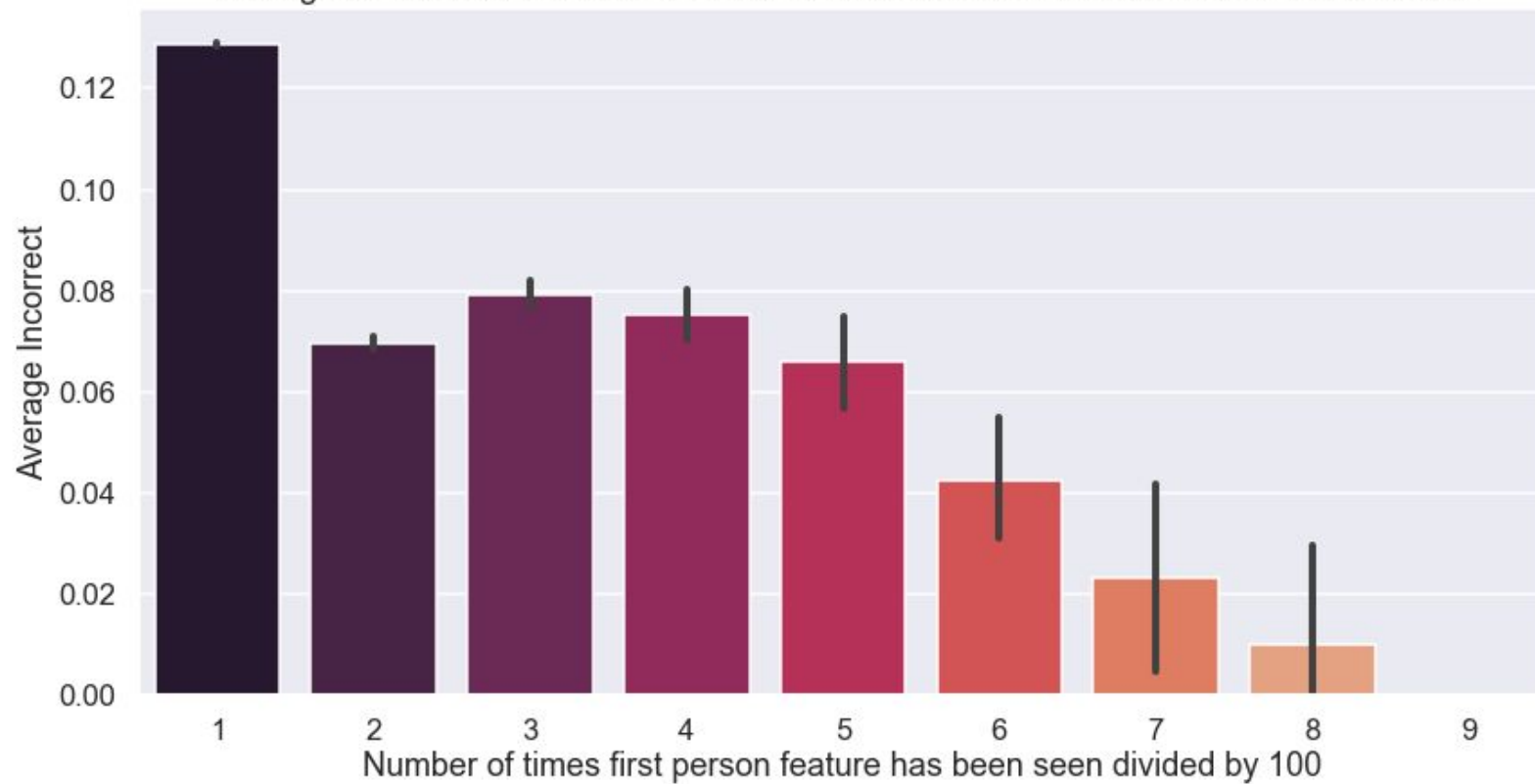
Feature Engineering

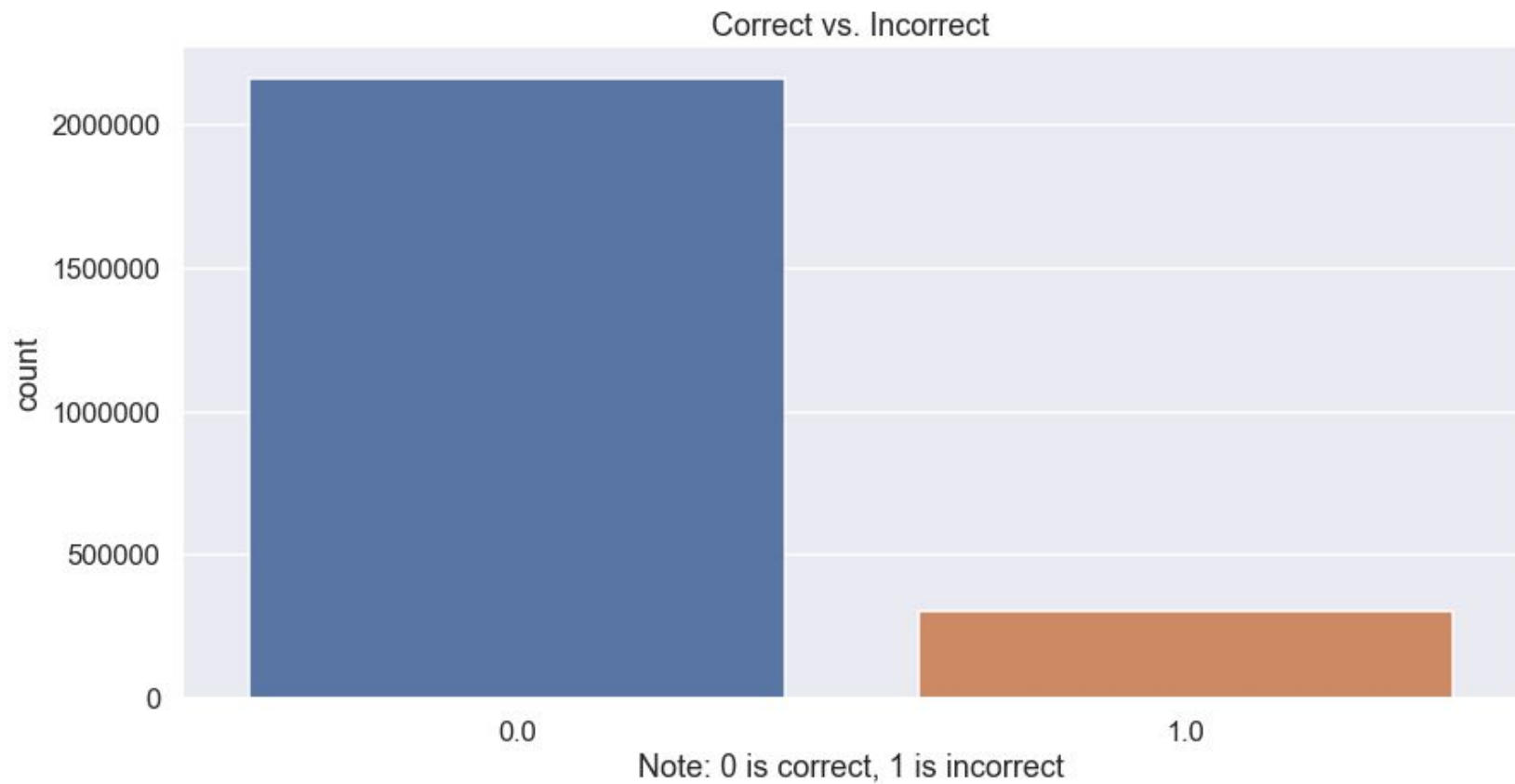
- English:
 - Is English (the target language) a *de facto* language in at least one of the countries the user has used the app in?
- Morphological features and part of speech:
 - Is this the user's first, second, third, etc. time seeing this particular feature or part of speech

Days user had used Duolingo



Average Incorrect vs Number of Times User Sees a Word with the First Person Feature





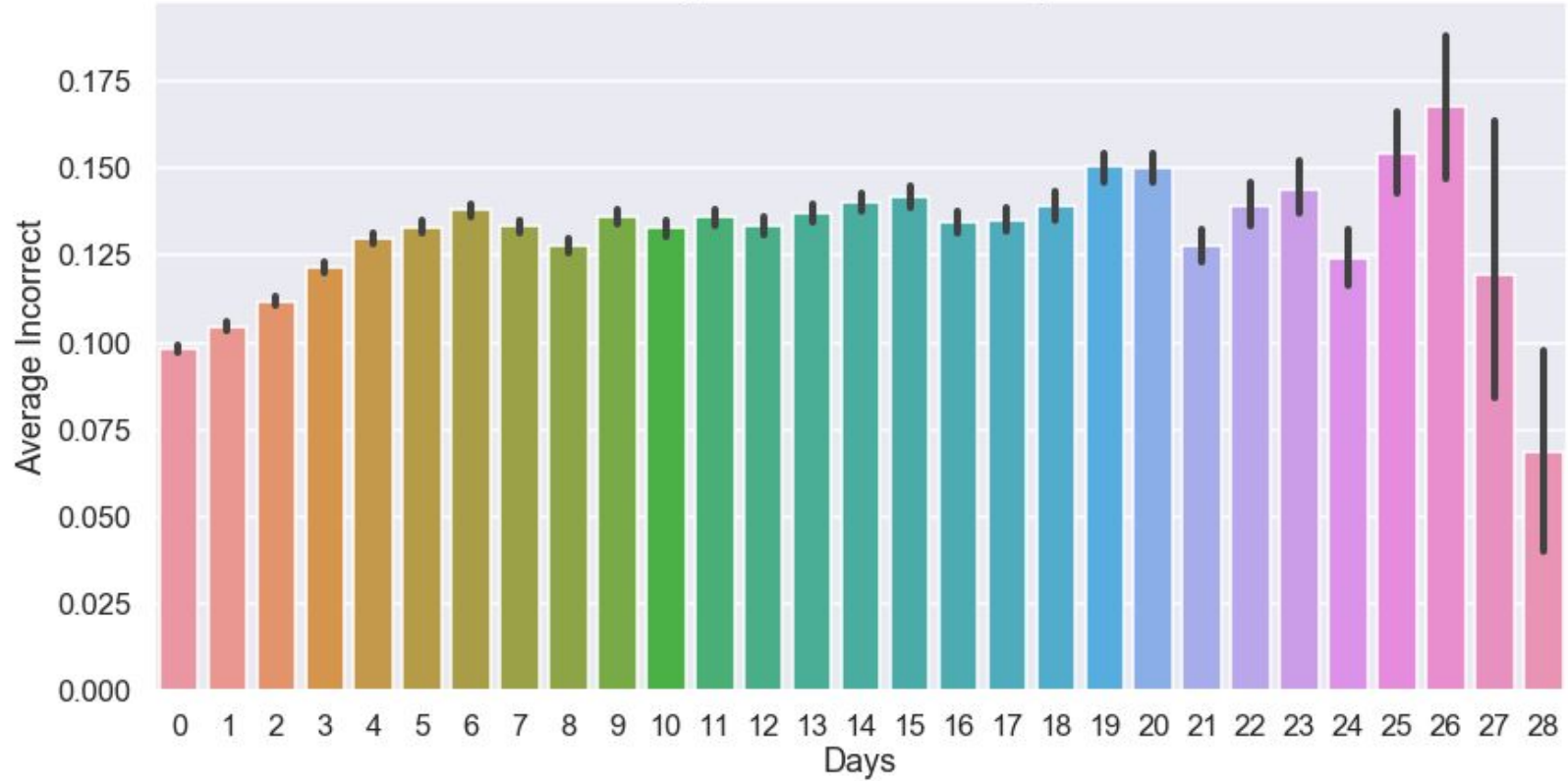
Addressing potential issues:

- Dimensionality
 - Final data frame has 58 features
 - Principal Component Analysis
 - Maintain 85% of variance: 36 features
- Imbalanced Data
 - Oversampling: SMOTE (Synthetic minority oversampling technology)
 - This greatly increased the F1 value (0.002 -> 0.29)

Results:

- Utilized Random Forest Model
- Accuracy: correctly predicted observations/total observations (0.53)
- F1: keeps a balance between Precision and Recall. (0.29)
 - Useful for data with uneven class distribution
 - Model has a good recall (0.776), but is imprecise (0.18)
 - Selects a lot of relevant items, but not all relevant items are selected
- AUC ROC: False Positives vs True Positives (0.64)

Days user had used Duolingo



Conclusion

- Modeling can help developers maintain a consistent difficulty level
 - Maintaining this difficulty level can help with user retention
- Users, primarily at the beginning of using the app, can be less predictable
 - Collect data for users who have been using the app for longer
- Whether or not the user was in an English speaking country seemed to be important
 - Tailor user content by location



5 in a row! Well done!



Your family is safe today!

Thank you!

```
[123]: forest = RandomForestClassifier(n_estimators=100, criterion='gini', max_depth= 5, class_weight="balanced")

[124]: fit_model_and_print_metrics(stk_indices, X_res, y_res, forest, print_metrics=True)

accuracy_score: 0.7623190628776095
precision_score: 0.7406164393044747
recall_score: 0.8048450742400465
f1_score: 0.7685893624664137
roc_auc_score: 0.7623190628776095

[124]: {'accuracy': 0.7623190628776095,
       'precision': 0.7406164393044747,
       'recall': 0.8048450742400465,
       'f1': 0.7685893624664137,
       'roc_auc': 0.7623190628776095}

[125]: fit_model_and_print_metrics(stk_indices_test, X_te, y_te, forest, print_metrics=True)

accuracy_score: 0.5349370472558669
precision_score: 0.18063739000198628
recall_score: 0.7761605271382962
f1_score: 0.29306054132113857
roc_auc_score: 0.638447716909507

[125]: {'accuracy': 0.5349370472558669,
       'precision': 0.18063739000198628,
       'recall': 0.7761605271382962,
       'f1': 0.29306054132113857,
       'roc_auc': 0.638447716909507}
```