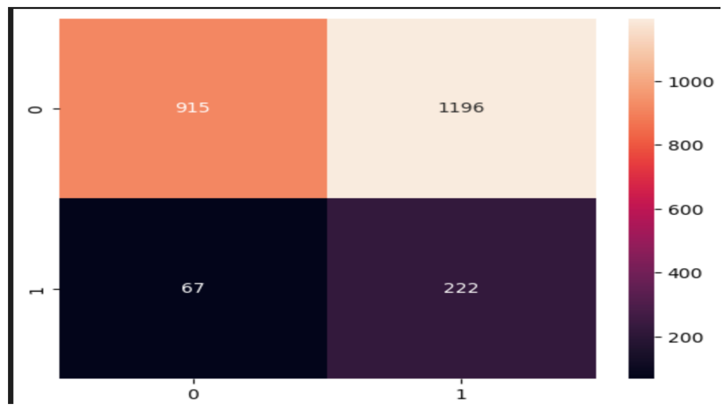
Given 12,000 rows of data for users who signed up for a product in the last 2 years, our goal is to identify which factors predict future user adoption.

Results:

Between the 3 models used for this analysis, **LightGB** was the best model. The results were validated using 5-fold cross validation.

	precision	recall	f1-score	support
0	0.93	0.43	0.59	2111
1	0.16	0.77	0.26	289
accuracy			0.47	2400
macro avg	0.54	0.60	0.43	2400
weighted avg	0.84	0.47	0.55	2400



- Of the 289 adopted users in the test set, the model is correctly identifying 77% of them (recall).
- Of the 2111 non-adopted users in the test test, 43% are correctly identified.
- The model predicted that 1418 users would be adopted users. Only 16% of those were correct (precision)
- The overall accuracy is 47%.

Important Features:

- **`month_created`**: From the EDA we can see that users that created their accounts between June and November had a higher rate of adoption than the rest of the year. June and October had the highest with approximately 16% of users being an adopted user
- **`creation_source_PERSONAL_PROJECTS`**: Only 7% of users who created their account for personal projects end up becoming an adopted user.
- **`org_type_uncommon`**: 15% of users from the uncommon org group are adopted users, whereas only 11% from the common group are.

Conclusion and Recommendations:

The results of this analysis are less than ideal, but there are still some insights to be gained from it. The time of year, the creation source, and organization type are decent predictors of whether or not someone will be an adopted user. Since less users are adopting the product from December until May, maybe implementing a discount for those months can encourage user retainment. Users who are creating an account for personal projects seem to not be happy with the product. I would suggest implementing a survey to these users to see what is missing from the product that would help retain them. Lastly, it could be helpful to have more information about the specific organizations. Without context for what the actual product is it is hard to give specifics, but for instance if users from a tech organization use the product more than users from an education organization, that can be useful for making future predictions. Overall, more user data needs to be available in order to increase the prediction accuracy of the model.