Reiss Pikett
Megan Stevens

Classification of Conversation Audio Files

Project Goals

Privacy is a big issue when discussing finances and to ensure motor finance companies are not divulging private information to strangers, phone calls must meet security regulatory compliance standards.

Trackback has recently taken on the task of measuring the volume of calls made by clients which meet these standards. Currently, the principle way of doing this is to manually listen to and check the audio call recordings. Our aim for this project was to develop a system to perform this check automatically, using voice to text transcription services and Machine Learning algorithms in application with natural language processing. This essentially split the project into two main tasks: creating a voice to text engine benchmark that would to determine the best transcription service, and training a machine learning algorithm to read the transcription and determine a call's level of compliance.

Voice to Text Software

Accuracy of speech to text engines
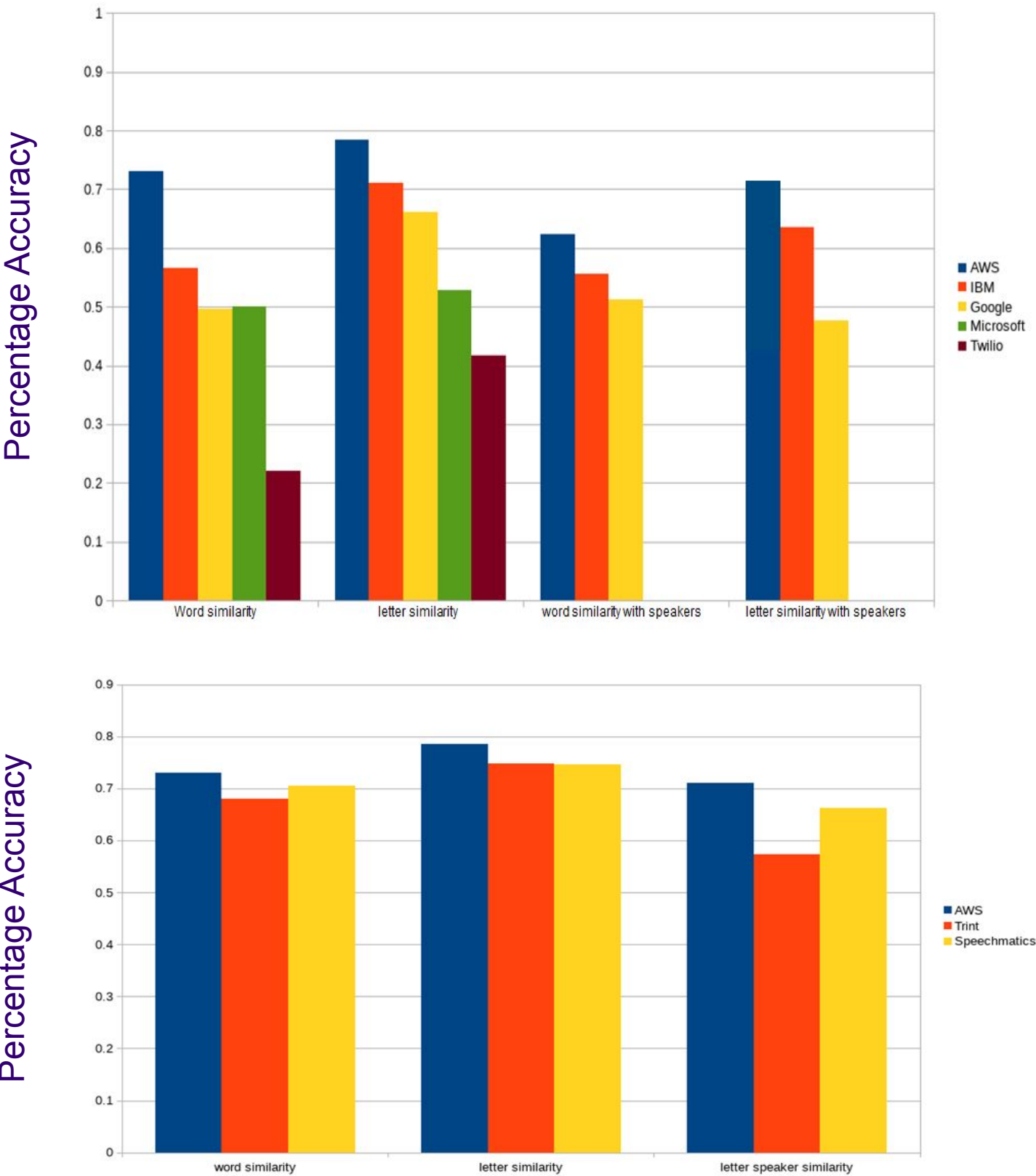
Most large tech companies, as well as a handful of less-known businesses, offer a speech to text software. The main engines that we investigated were Amazon Web Services Speech to Text, IBM Watson, Google Speech Recognition, Microsoft Azure, Twilio, Speechmatics, and Trint.

When comparing the different options we considered three main factors:

 Ease of set-up
 Transcription Accuracy
 Accuracy of Speaker Differentiation

AWS proved to be the most consistent at accurately transcribing recorded calls. While not being the easiest engine to set-up it was by no means too difficult.
Therefore we continued to use AWS as our voice to text engine as we entered the next phase of the project.

Summary of Tasks

> Investigate ways to classify audio files of conversations to determine if certain types of key phrases were spoken.

> Determine available transcription services, key phrases, and required criteria: e.g. accuracy, speed, regional speech, slang, etc.

> Create transcription service benchmark with respect to required criteria. Select most appropriate.

> Train Machine Learning Text Classification model to search for key phrases.

> Determine most effective method of applying algorithm to text.

> Create code to perform all key steps in one: Transcribe, Train, Classify

> Investigate methods of automated re-training of model with new data.

Machine Learning Algorithms

Compliant phone calls must include the caller mentioned that the call is being recorded and must include the caller confirming the identity of the customer by asking for three specific pieces of information.

Amazon offers a machine learning text classification algorithm called SageMaker BlazingText. It reads sentences and assigns them a label determined by the data it was trained with. This was perfect for our need case as we could feed sentences from the call transcription into the algorithm and have it output which of the four criteria the sentence satisfied or if it was simply a non-compliant sentence. If all four criteria were satisfied somewhere in the transcription then the call would be logged as compliant.
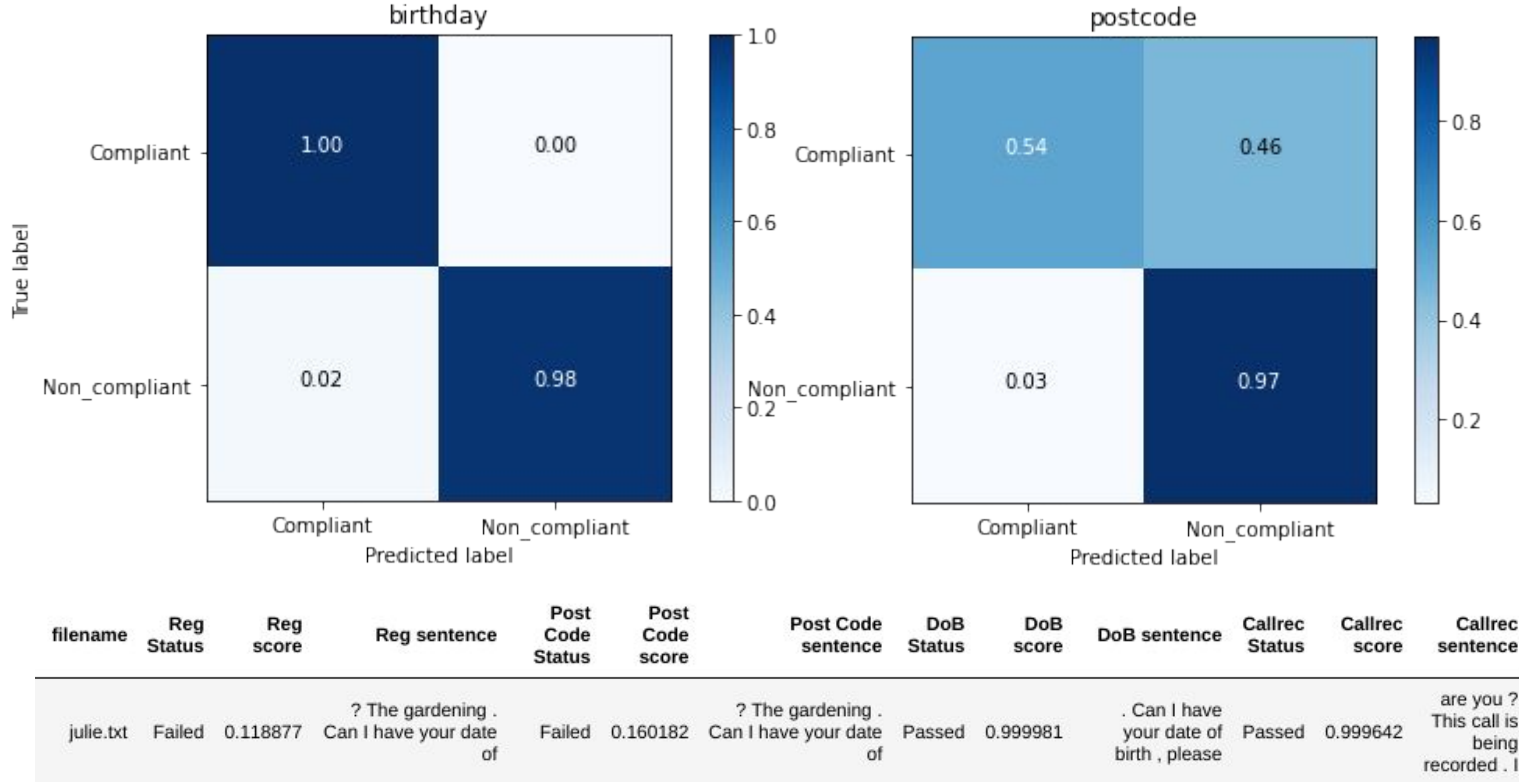
Two methods were used when training the algorithm. The first method looked for key words, while the second method also looked at the sentence structure of these compliant sentences.

Percentage accuracy of machine learning algorithm – Method 1
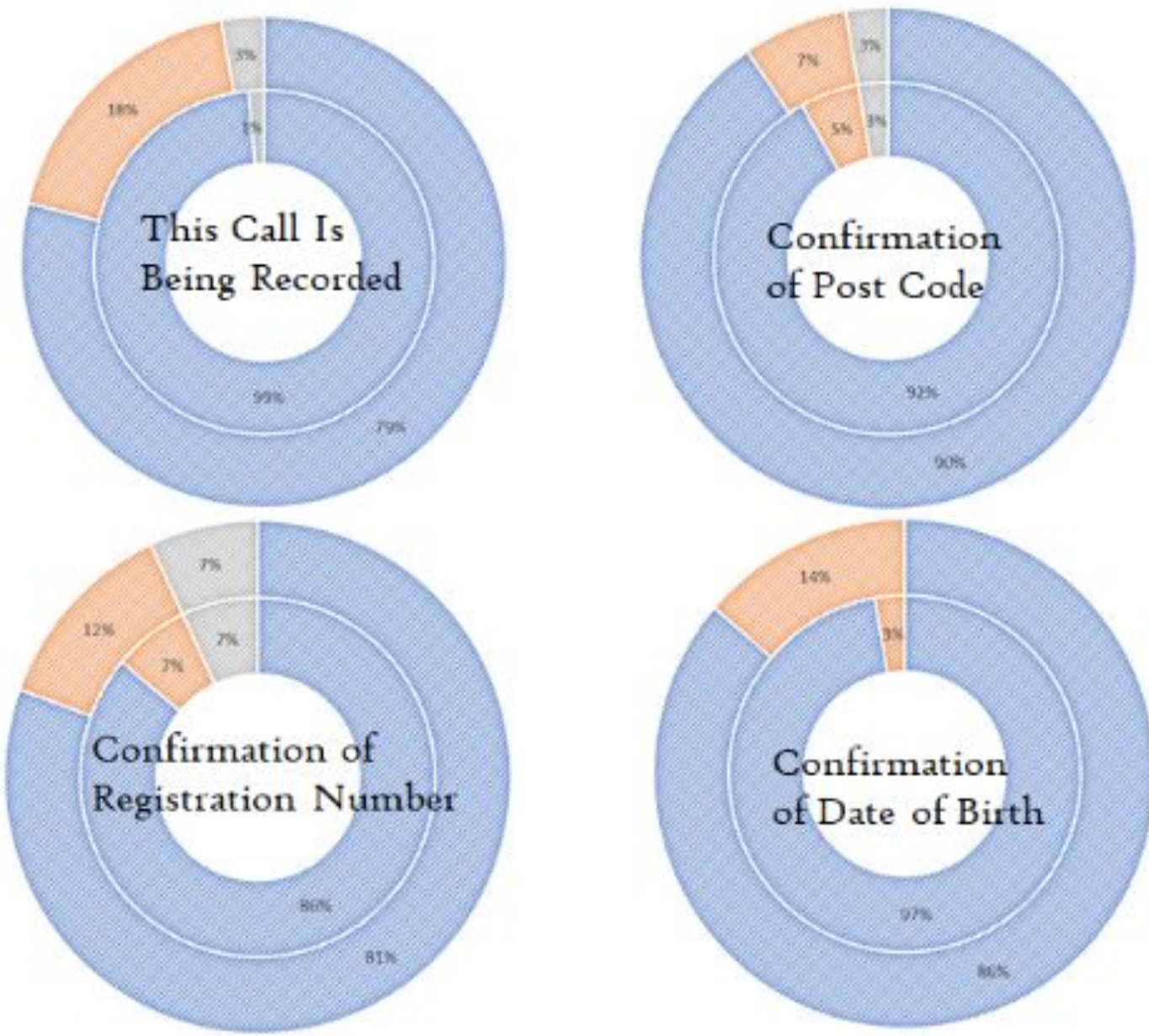
Accuracy by criteria – Method 2

Top: Confusion Matrix of labels sample, comparing the predicted compliance score vs the true, manually checked compliance score. Bottom: Sample output of classification and highest scoring sentence for each label

| filename	Reg Status	Reg score	Reg sentence	Post Code Status	Post Code score	Post Code sentence	DoB Status	DoB score	DoB sentence	Callrec Status	Callrec score	Callrec sentence
julie.txt	Failed	0.118877	? The gardening . Can I have your date of	Failed	0.160182	? The gardening . Can I have your date of	Passed	0.999981	. Can I have your date of birth , please	Passed	0.999642	are you ? This call is being recorded . I

The methods classify the transcribed text and analyse the results with respect to manually checked conversations. The output is the sentence with highest compliance likelihood for each label and the pass/fail status, which allows for continuous improvement of the model by manually checking and adding incorrectly classified sentences to their respective label training data.

Viability of use in production

The results of how accurately our methods were marking calls as compliant are encouraging and with tweaking to the models and retrained with more data this combination of Amazon programs should be very useful for the company in a production environment.

The transcription process takes significantly more time than the compliance checks however, even on a massive scale, the time taken to complete the whole process is viable. It is understood that most transcription services take half as much time as the length of the audio clip to create a transcript. The graph below shows that this relationship is complicated but does ten to follow this rule.

Time taken for audio files to be transcribed and compliance checked
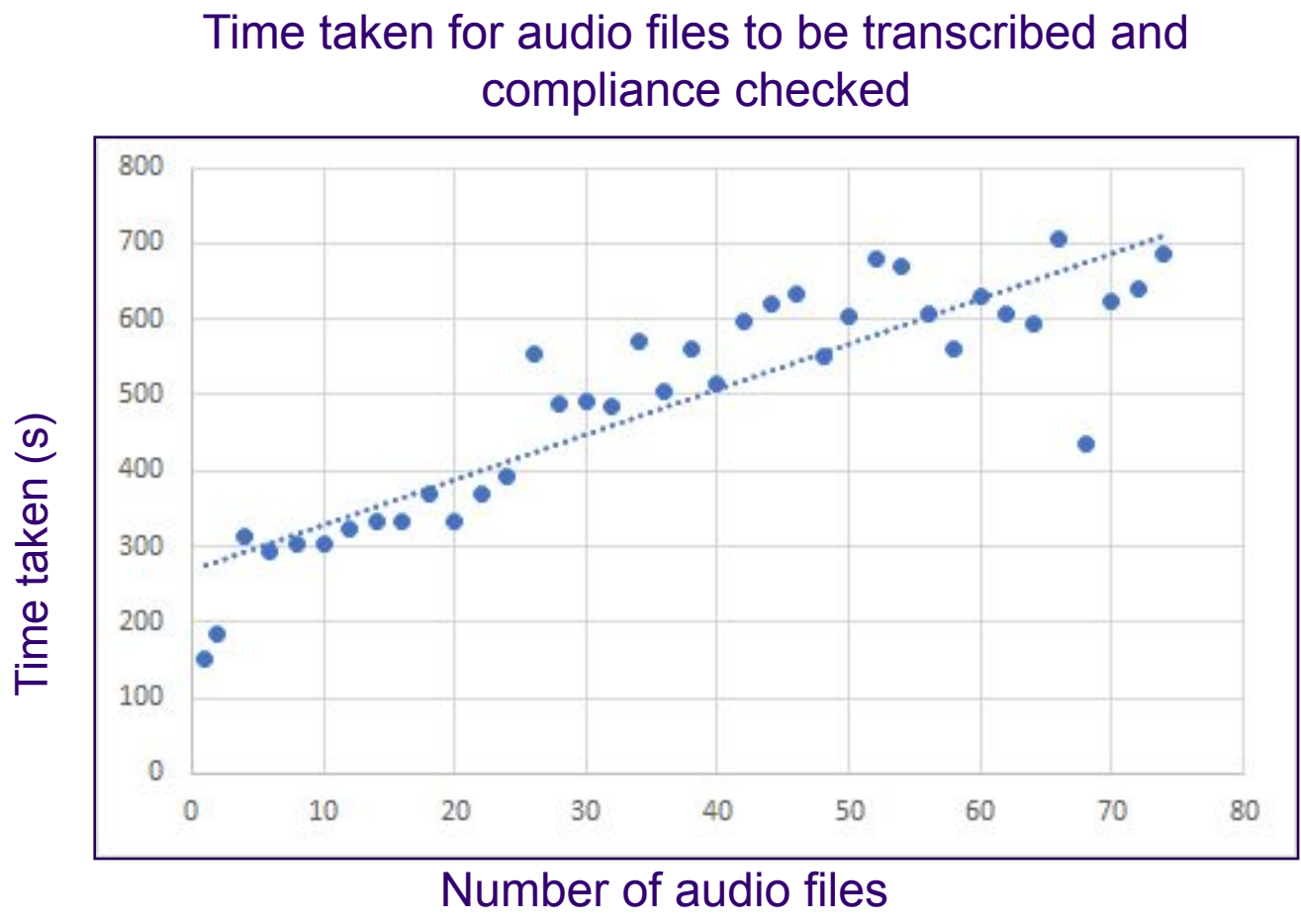
Possible Next Steps

This project was started in the hopes of being able to automate the process of checking conversations for compliance, but it has many other promising possibilities.

 The speech to text algorithm may be improved using the custom dictionary. This is a function that allows the addition of slang terms that may be used, and allows the transcription engine to favour key words over other similar sounding words.
 It would be very easy to train another BlazingText model with sentences that allow the conversations to be categorised in different ways, e.g. purpose of call.
 Implement methods of further tuning to train an algorithm by not only finding compliance questions but also to take note of answers and add corrections to the training data for ret-training.
 The text classification methodology could also be applied to verifying the compliance of e-mails.