



# **Bon Voyage!** **The Stress-Free,** **Automated Message** **Explainer (SAME)**

Let the robots do the *same ol'* work  
so you don't have to!

# Contents



## 01

### Introduction

What's the SAME &  
what's the difference?

## 02

### Data

### Exploration

What data was SAME  
trained on?

## 03

### Modeling

What's the brain  
behind SAME?

## 04

### Conclusion

How can we use it?

# There are so many messages in my inbox!

The SIA<sup>1</sup> Customer Service Centre receives thousands of feedback, compliments and complaints every week



# Help me sort out the enquiries!



**KrisFlyer &  
PPS Club (KF)**



**Lounge, Catering,  
Amenity (LCA) kits**



**Others**

Top 2 most-asked topics

**Objectives:** SIA management hired us to

- (1) Develop a predictive model to **automatically sort enquiries into the 3 topics**, and
- (2) Highlight the **frequently mentioned words** in KF and LCA



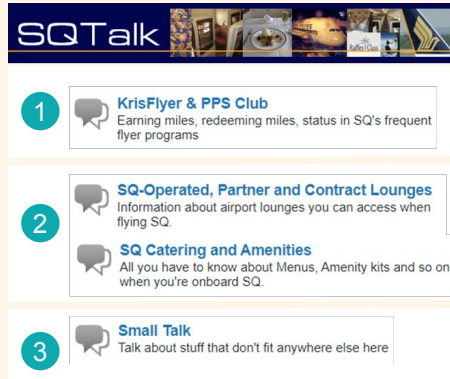
# **SAME:** **Always on the move**

SAME is a natural language processing model: it can do the same enquiry sorting as we can, but around the clock!

# The Approach



## Data extraction



**16k (40%), 10k (24%), 14k (36%) data rows respectively**



## Modelling

Try different **pre-processing, vectoriser, models** & pick the best combination



## Sentiment Analysis

Find the **top comments of KF and LCA**, and **analyse the sentiments** of these comments to derive insights

# Strap in! Diving into the data

The screenshot shows the SQTalk forum interface. The left sidebar contains navigation links for Forums, Blogs, Articles, and Groups, along with a list of topics. A red box highlights the thread title "Sticky: Qualifying as EG for the first time" in the Topics section. A red line points from this box to a detailed view of the thread on the right. The thread is titled "Sticky: Qualifying as EG for the first time" and is started by Vtac82. The thread content includes a quote from Nick C, a response from Vtac82, and a response from SuperJonJon. The thread is marked as a sticky post.

**SQTalk**

Forums Blogs Articles Groups

Today's Posts Member List Calendar

Forum > Singapore Airlines > KrisFlyer & PPS Club

**KrisFlyer & PPS Club**

Earning miles, redeeming miles, status in SQ's frequent flyer programs

TOPICS LATEST ACTIVITY MY SUBSCRIPTIONS PHOTOS

Topics

- Sticky: Qualifying as EG for the first time  
Started by Vtac82, 14 May 2008, 14:32 PM
- Sticky: Which FFP for me? Master Discussion  
Started by YK@HKG, 25 June 2007, 06:46 PM
- #SQMileageTeam  
Started by SQ218, 28 November 2022, 05:54 PM
- Advice sought - Changing redemption bookings  
Started by Jalin, 24 October 2022, 08:44 PM
- First Savers SYD-SIN  
Started by rbarberini, 24 September 2022, 10:40 AM
- Waitlist on Redemption booking  
Started by SQ218, 18 May 2022, 06:51 PM
- request to cancel miles (FQTV and FQTS)  
Started by Stripakis, 25 July 2022, 08:12 PM
- Redemption requirements up 10%  
Started by alimotor\_2000, 7 June 2022, 10:53 PM
- Missing Krisflyer Miles for PPS and PPS Solitaire  
Started by 11A, 16 March 2022, 10:38 PM
- Mixed SQ/Star Alliance Award  
Started by SQ747, 21 February 2022, 06:42 PM
- Contact Centre Unable to See Seats That Are Available  
Started by SQ747, 21 February 2022, 06:30 PM
- F Advantage LHR to SYD return help  
Started by VC10 Boy, 30 December 2021, 07:15 PM

14 May 2008, 05:43 PM #4

Originally posted by **Nick C**

There's no need to panic. Been in a similar situation before. What you see on KF is based on what you need to accumulate to renew your KF membership thru April 2009. The system tracks your mileage accurately, you will be upgraded to KF-G after your upcoming trip.

Yeah? What was concerning was that even the lady on the phone told me I had to collect further 48,064 to get KF G.

Anyway, if that's the case, does that mean by transit in Singapore I can get Gold Lounge access?

From experience, flying with SQ, miles are credited immediately upon departure, that means that by the time I transit in Singapore, I'll be Gold, wonder if I can gain access to Gold lounge.

Either way, I'll keep you guys posted on Sunday.

14 May 2008, 05:44 PM #5

Firstly, Welcome to SQTalk. Vtac82!

Krisflyer miles and Elite miles are actually two separate things. Elite miles are earned only thru flying but Krisflyer miles can be earned through credit cards, hotel stays, etc.

I am guessing that you have flown about 26,936 miles or slightly more. When you hit 25K or slightly above, you achieved KF Silver in May thus the counter has reset. You have flown 1,936 miles since achieving Silver and would require another 23,064 miles to get Gold.

I hope that I am wrong and you get your KF Gold this Sunday.

14 May 2008, 05:46 PM #6

Originally posted by **SuperJonJon**

Firstly, Welcome to SQTalk. Vtac82!

Krisflyer miles and Elite miles are actually two separate things. Elite miles are earned only thru flying but Krisflyer miles can be earned through credit cards, hotel stays, etc.

I am guessing that you have flown about 26,936 miles or slightly more. When you hit 25K or slightly above, you achieved KF Silver in May thus the counter has reset. You have flown 1,936 miles since achieving Silver and would require another 23,064 miles to get Gold.

I hope that I am wrong and you get your KF Gold this Sunday.

Tier bonuses also does not count towards Elite qualification / requalifications.

1 2 3 4 11 51 101 130 Next

1. Properly spelled words
2. Absence of Singlish
3. Replies to comments will duplicate words
4. Thread titles are important as they contain many key words

# What's the Model Answer?

Bolded – different from baseline

Steps	Baseline	Alternate 1	Alternate 2	Alternate 3
Pre-processing	<ul style="list-style-type: none"><li>Basic data cleaning</li></ul>	<ul style="list-style-type: none"><li>Basic data Cleaning</li><li><b>Remove duplicated sentences</b></li></ul>	<ul style="list-style-type: none"><li>Basic data Cleaning</li><li><b>Remove duplicated sentences</b></li></ul>	<ul style="list-style-type: none"><li>Basic data Cleaning</li><li><b>Remove duplicated sentences</b></li></ul>
Vectorisation	CountVectoriser	CountVectoriser	<b>TFIDF</b>	<b>WordVectoriser</b>
Model	Multi-Naïve Bayes	Multi-Naïve Bayes	Multi-Naïve Bayes	Multi-Naïve Bayes

Best performing pre-processing & vectorisation combination

2

Modelling trials:

Model	Multi-Naïve Bayes	<b>Random Forest</b>	<b>XGBoost</b>	<b>SVM</b>
-------	-------------------	----------------------	----------------	------------

**Weighted-average ROC AUC & weighted-average f1 (KF, LCA) scores used to evaluate model performance**

\* The alternate 1, 2, 3 also had 4 additional stopwords (iirc, imo, imho, btw) extracted from a "SQTalk Abbreviations, Slangs, Definitions, Phrases" thread. This had little effect on the model performance.



# Preprocess, Vectoriser Comparison

	Baseline Model	Alternate 1 *Best performance*	Alternate 2	Alternate 3
Pre-processing	- Basic cleaning	- Basic cleaning	- Basic cleaning - Remove duplicated sentences	- Basic cleaning - Remove duplicated sentences
Vectoriser	CountVectoriser	TFIDF	CountVectoriser	TFIDF
Model	Multinomial Naive Bayes	Multinomial Naive Bayes	Multinomial Naive Bayes	Multinomial Naive Bayes
Weighted-average ROC AUC	0.877	0.889	0.790	0.801
Weighted-average f1-score (kf, lca)	0.743	0.748	0.623	0.628

Best performance: Basic clean, TFIDF

## Remarks:

- TFIDF had slightly better performance than CountVectoriser
- **Removing duplicated comments seemed to have an adverse effect** on model performance; this suggests that the comments that people reply to usually have crucial key words in them

# Model Comparison

	Model 1 (from notebook 3)	Model 2	Model 3	Model 4 *Best performance*
Vectoriser	TFIDF	TFIDF	TFIDF	TFIDF
Model	Multinomial Naive Bayes	Random Forest	XGBoost	SVM
Weighted-average ROC AUC	0.889	0.872	0.913	0.913
Weighted-average f1-score (kf, lca)	0.748	0.721	0.790	0.802

Best performance: SVM

Remarks:

- **Model 3 performed the best** with the following params:
  - TfidfVectorizer(max\_features=500, stop\_words='english')
  - SVC(C=1, gamma=1, probability=True, random\_state=42)
- SVM edged out due to its **slightly better f1-score**

# Sentiment Analysis: Top Words

## KrisFlyer & PPS Club

	count
miles	8899
SQ	8627
posted	7734
Originally	7367
PPS	5771
SIN	4730
flight	3692
KF	3616
just	3547
year	2822

## Lounges, Catering, Amenity Kits (LCA)

	count
SIN	7319
lounge	4650
SQ	4337
posted	3819
Originally	3637
The	2213
served	2031
chicken	1928
flight	1872
vegetables	1805



1. Find top words
2. Find sentences that contain these words
3. Sort the sentiment score by:

Sentiment	Score	Assigned value
Very Negative	Under -0.5	1
Negative	Between -0.5 and -0.1	2
Neutral	Between -0.1 and 0.1	3
Positive	Between 0.1 and 0.5	4
Very Positive	Over 0.5	5

# Sentiment Analysis: Results

Sentiment Score	KrisFlyer 'miles'	KrisFlyer 'PPS'	KrisFlyer 'KF'	LCA 'lounge'	LCA 'SQ'	LCA 'served'
Very negative (1)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Negative (2)	<0.1	<0.1	<0.1	<0.1	0.1	0.1
Neutral (3)	0.3	0.4	0.4	0.4	0.4	0.6
Positive (4)	0.6	0.6	0.6	0.5	0.4	0.3
Very positive (5)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Overall:	Positive	Positive	Positive	Positive	Positive	Positive

\* Due to rounding, total may not add up to 1

## Remarks:

- Sentiments for comments containing top words are **positive** → **these are areas of strength for SIA**
- Separate deep-dive into these comments can be conducted to find out the reasons why (e.g. good service, good exclusive deals for members, good food served etc.)

# Conclusion



## Model

**Outcome: Successfully developed SAME:** a prediction model with weight-average ROC AUC (0.913) and f1-scores (0.802)



## Sentiment Analysis

Comments with 'miles', 'PPS', 'KF', 'lounge', 'SQ', 'served' have **largely positive sentiments** → **areas of strength for SIA**



## Next Steps (areas for future improvement)

- **Sort by topic (e.g. KF, LCA, others):** SAME can be used as a backend engine for a chatbot to sort incoming queries
- **Sort by type (feedback, complaints, compliments):** SAME + Sentiment Analysis
- **Incorporate Singlish** into sentiment analysis to improve performance
- Find **top words with negative sentiments** to identify and improve on areas of weaknesses

**Thanks!**

**Your work will never  
be the *same* again**



# Annex A - Weighted Average f1-Score

Label	Per-Class F1 Score	Support	Support Proportion	Weighted Average F1 Score
 Airplane	0.67	3	0.3	$(0.67 * 0.3) +$ $(0.40 * 0.1) +$ $(0.67 * 0.6)$ <b>= 0.64</b>
 Boat	0.40	1	0.1	
 Car	0.67	6	0.6	
Total	-	10	1.0	

Support refers to the number of actual occurrences of the class in the dataset.

For example, the support value of 1 in Boat means that there is only one observation with an actual label of Boat. ([source](#))