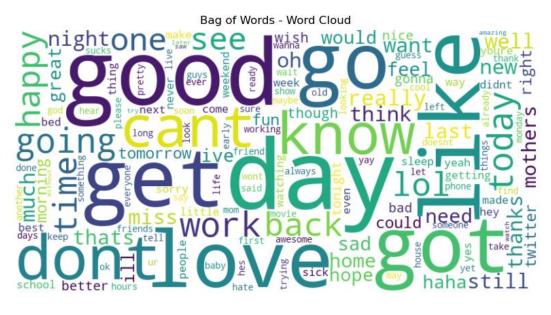# SENTIMENT ANALYSIS

Name: *T Ramsrikumar*

Reg No: *21011101096*

Pipeline

- *Load Dataset*
- *Clean the Dataset*
- *Text Pre-Processing*
- *Text Vectorization*
- *Build and Train Model*
- *Plot Classification Report*

Frequent Words

Bag of Words - Word Cloud

Output:

```
What is not to like about this product.  :  negative
Not bad.  :  negative
Not an issue.  :  negative
Not buggy.  :  neutral
Not happy.  :  positive
Not user-friendly.  :  negative
Not good.  :  positive
Is it any good?  :  positive
I do not dislike horror movies.  :  negative
Disliking horror movies is not uncommon.  :  negative
Sometimes I really hate the show.  :  negative
I love having to wait two months for the next series to come out!  :  positive
The final episode was surprising with a terrible twist at the end.  :  neutral
The film was easy to watch but I would not recommend it to my friends.  :  neutral
I LOL'd at the end of the cake scene.  :  neutral
```

Accuracy Scores:

Bag of words	-	*0.65*
TF – IDF	-	*0.63*
Continuous Bag of words	-	*0.54*
Skip gram	-	*0.52*
Word2Vec – Twitter Glove model	-	*0.63*

Limitations of each Vectorizing Technique

- **Bag of Words (BOW)** : *Ignores word order, leading to a loss of sequential information. High-dimensional vector representation can be computationally inefficient.*

- **TF-IDF** : *Doesn't capture word semantics or relationships, treating each word independently. Importance is based on frequency and rarity, ignoring context.*

- **Continuous Bag of Words (CBOW)** : *Doesn't consider word order, similar to BOW. Fixed context window may not capture long-range dependencies effectively.*

- **Skip-gram** : *Computationally expensive, especially for large vocabularies. Requires more data for effective training compared to CBOW.*

- **Word2Vec** : *May not perform well for rare words or infrequent terms. Fixed-size vectors might not capture nuances of word meanings in different contexts.*

Source Code: