



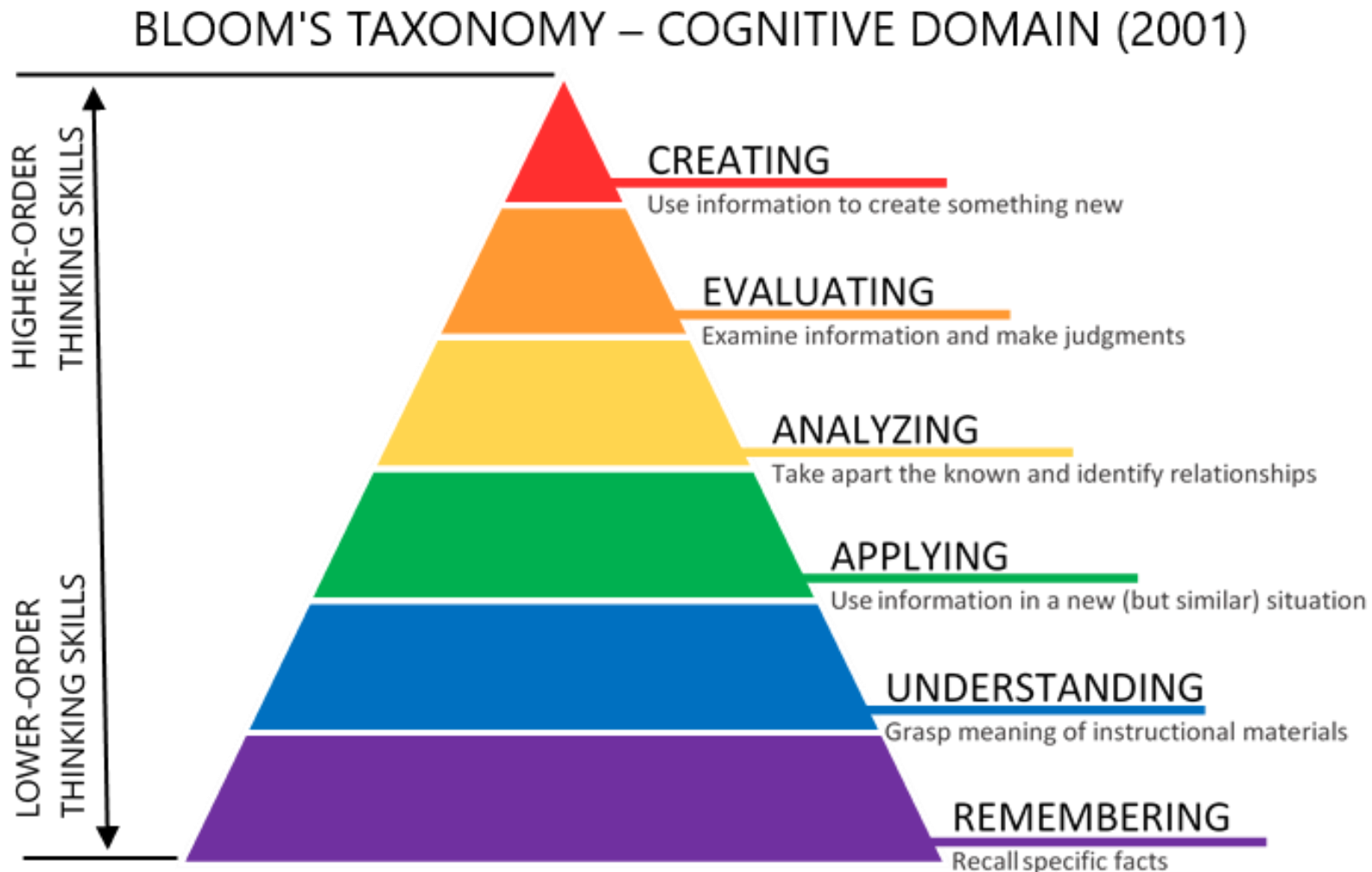
# Bloom's Taxonomy ML Classifier

Neha Vijay Patil

---

# Defining the problem

# Problem Definition



Facilitating the creation of balanced question papers to evaluate students on different levels based on Bloom's Taxonomy.

# Problem Definition

- Existing method of implementation: Simple mapping of verbs -> levels

## Limitations:

- action verbs are not always the indicator of level. (what is...)
- some verbs are associated with multiple levels. (describe)

## Proposed solution:

- ML Classifier will solve the above issues and automates the categorisation process

---

# Literature Survey

Papers	[11]	[12]	[13]	[9]	[15]	[16]	Our Algorithm
Feature extraction	NA	Chi-Square+Laplace Smoothing	Chi-Square,Mutual Information, and Odd Ratio	TFPOS-IDF and word2vec	NA	Pos Tagging and Word Vector	TFPOS-IDF
Classifiers	SVM, K-NN.	Naive Bayes	Naïve Bayes(NB), (KNN).	KNN, Logistic Regression, SVM	SVM-Light	Rule based grammar, , Ensemble Technique on KNN, SVC and NB, Wordnet.	Ensemble Technique on KNN, SVC and NB.
Accuracy	SVM: 69% K-NN: 65%	60.63%.	KNN: 87%. NB: 83%.	KNN: 71.1%, LR: 82.3% SVM: 83.7%	87.4%. Poor recall and F-measure values	Ensemble: 82.5% Rule Based: 60%	93.5% Precision: 94% Recall: 93%

---

# Data & Model Design

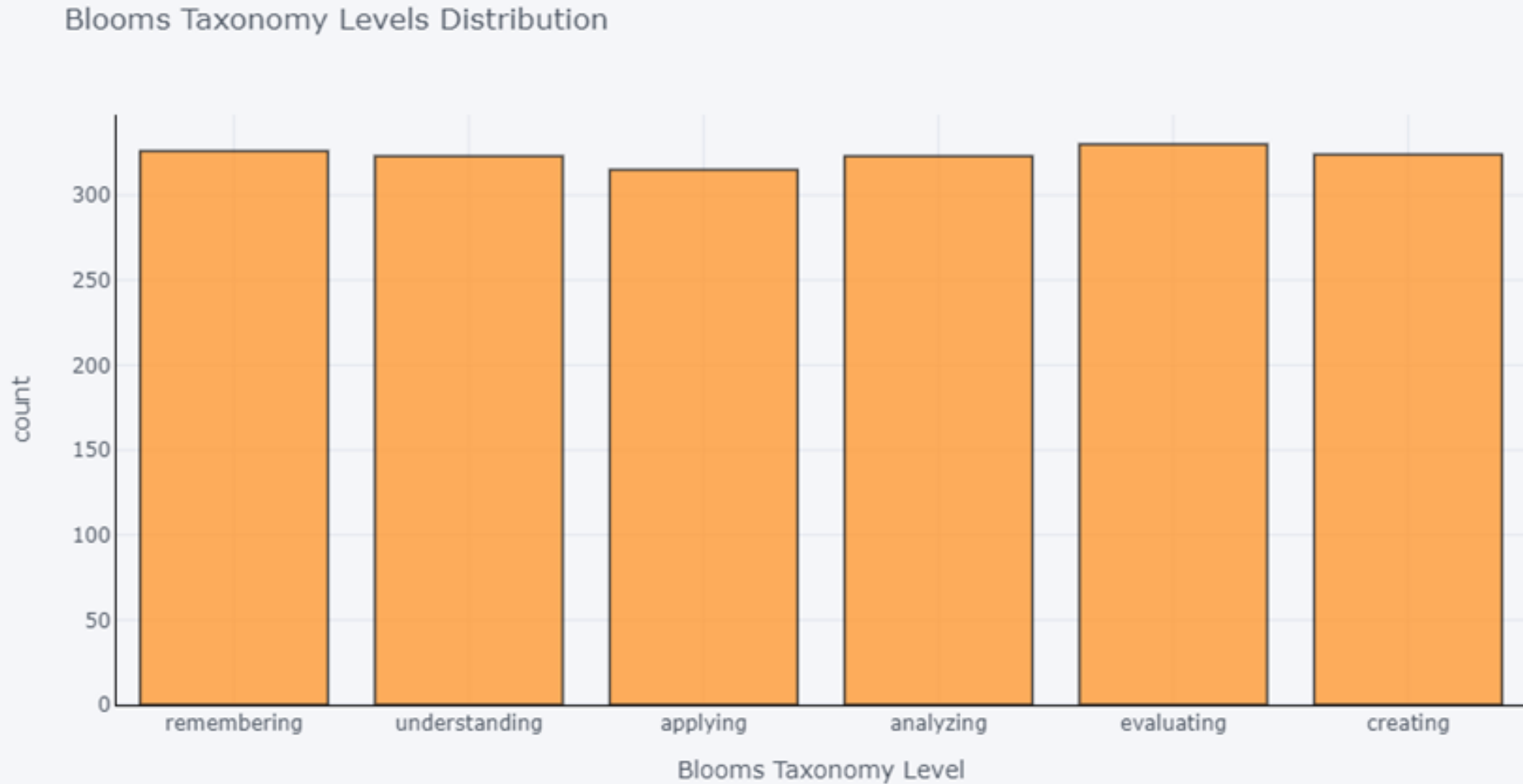
# Snippet of the dataset

4] :

	Text	Target
1221	name five cities in us	remembering
588	revise the story and specify the type of audie...	creating
302	analyze the movements and sounds of a frog	analyzing
37	list reserved words in c programming	remembering
281	what approach would you use to prove the truth...	applying
403	can you develop a proposal which would	evaluating
547	given the data weve looked at on this topic ev...	creating
530	determine if a character's actions were heroic	creating
332	examine what helps to make a good olympics thi...	analyzing
25	identify the correct definition of osmosis	remembering



# Distribution of levels in dataset

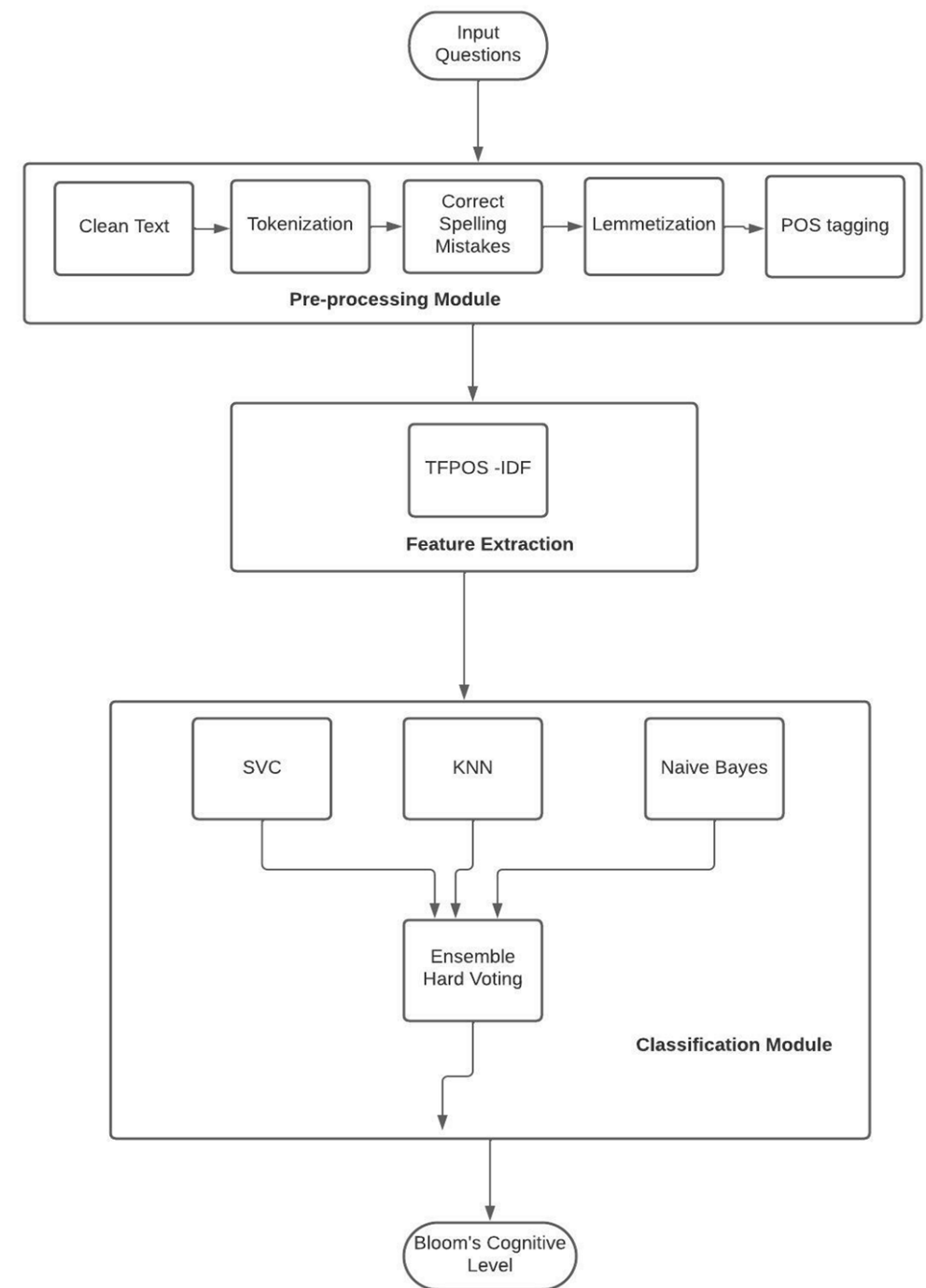


# Model Design

- Customised feature extraction algorithm to give high weightage based on positioning.

$$TFPOS = \frac{\text{Frequency of word in document} * \text{weight of the word}}{\sum \text{weights of each word in document}}$$
$$IDF = 1 + \log\left(\frac{\text{number of documents in collection}}{\text{number of documents which contains the word}}\right)$$
$$TFPOS-IDF = TFPOS * IDF$$

- Ensemble ML classifier combining results of 3 models: KNN, SVC and Naïve Bayes using hard voting.



---

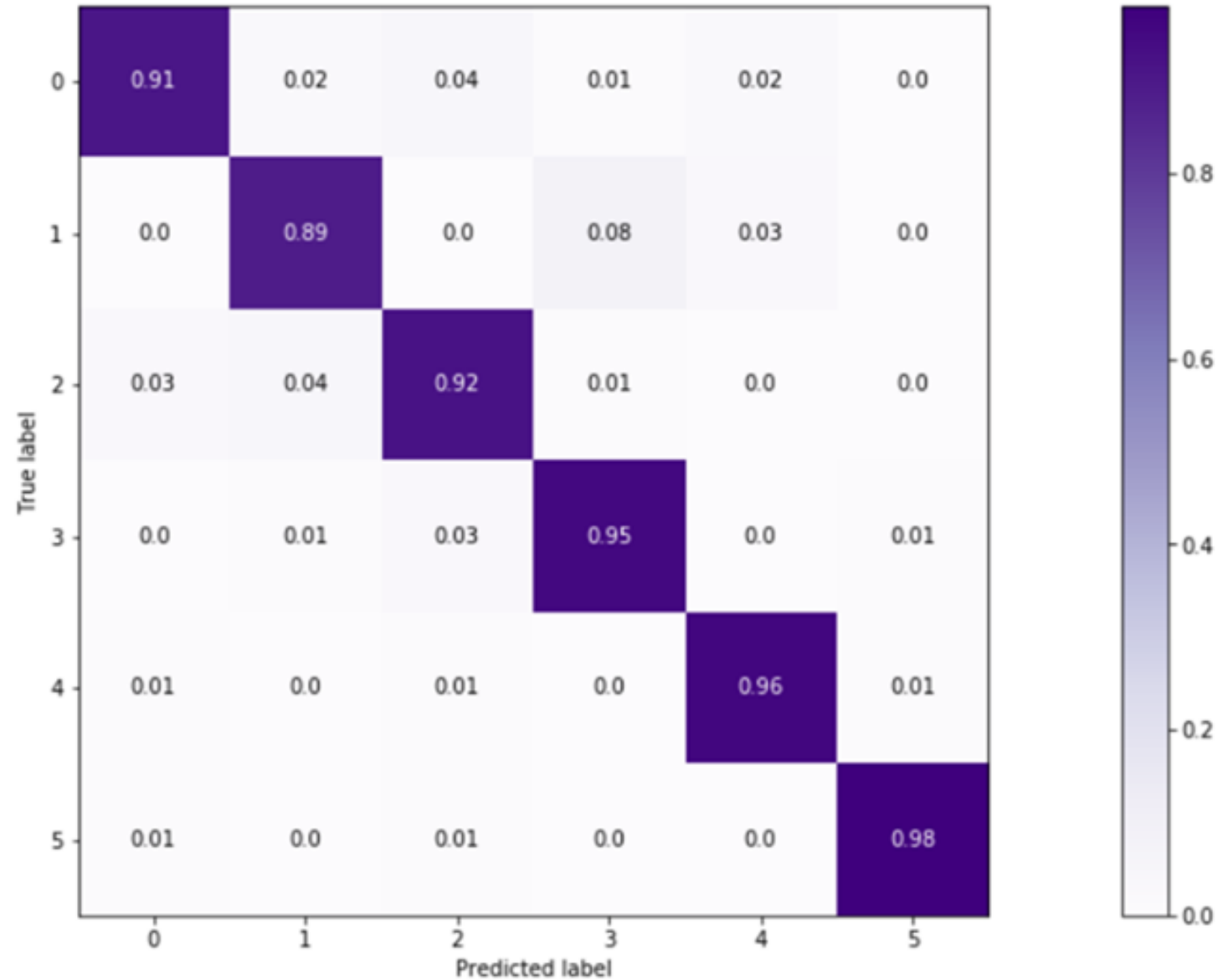
# Implementation

<https://github.com/neha7799/BloomsTaxonomyClassifier/blob/main/model%20presentation.ipynb>

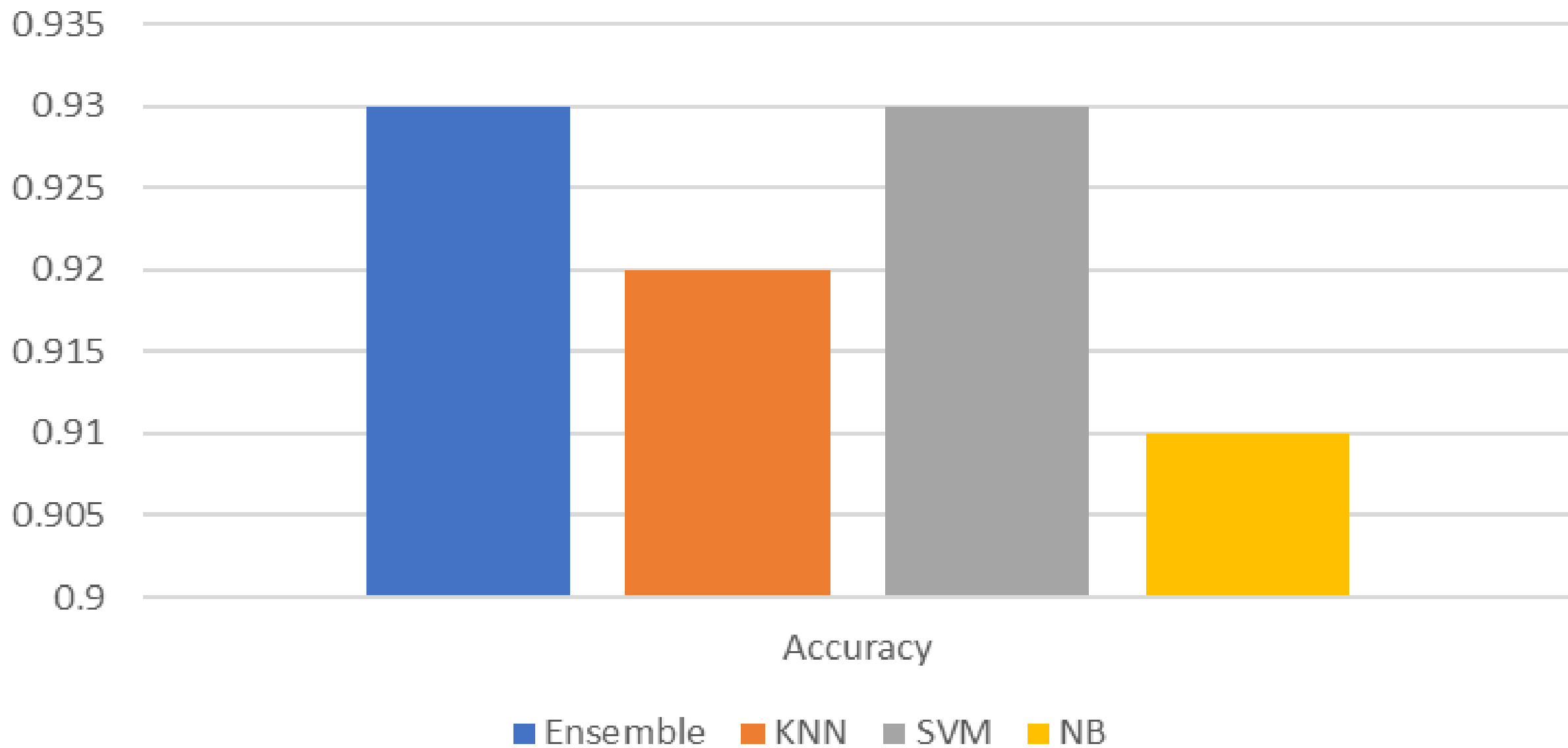
---

# Results

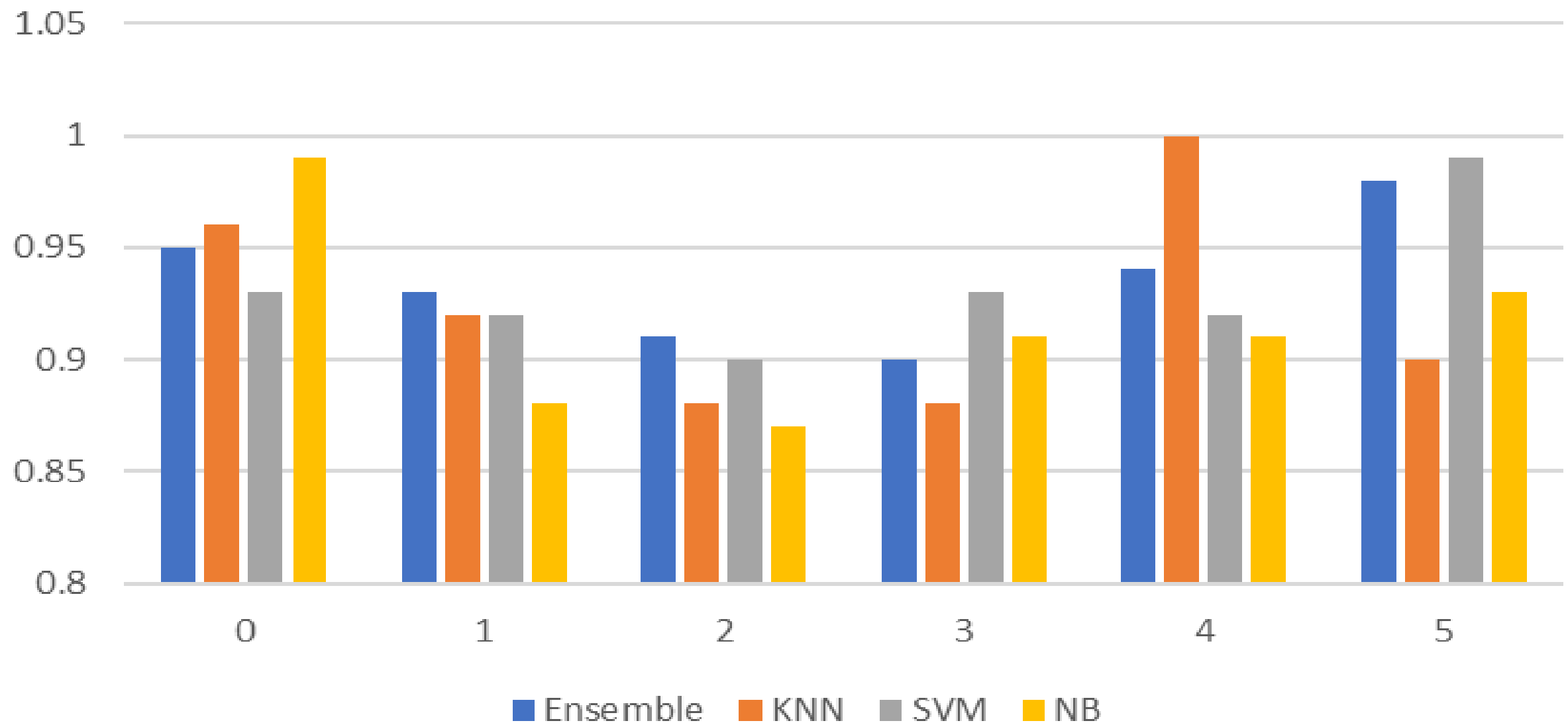
# Confusion Matrix



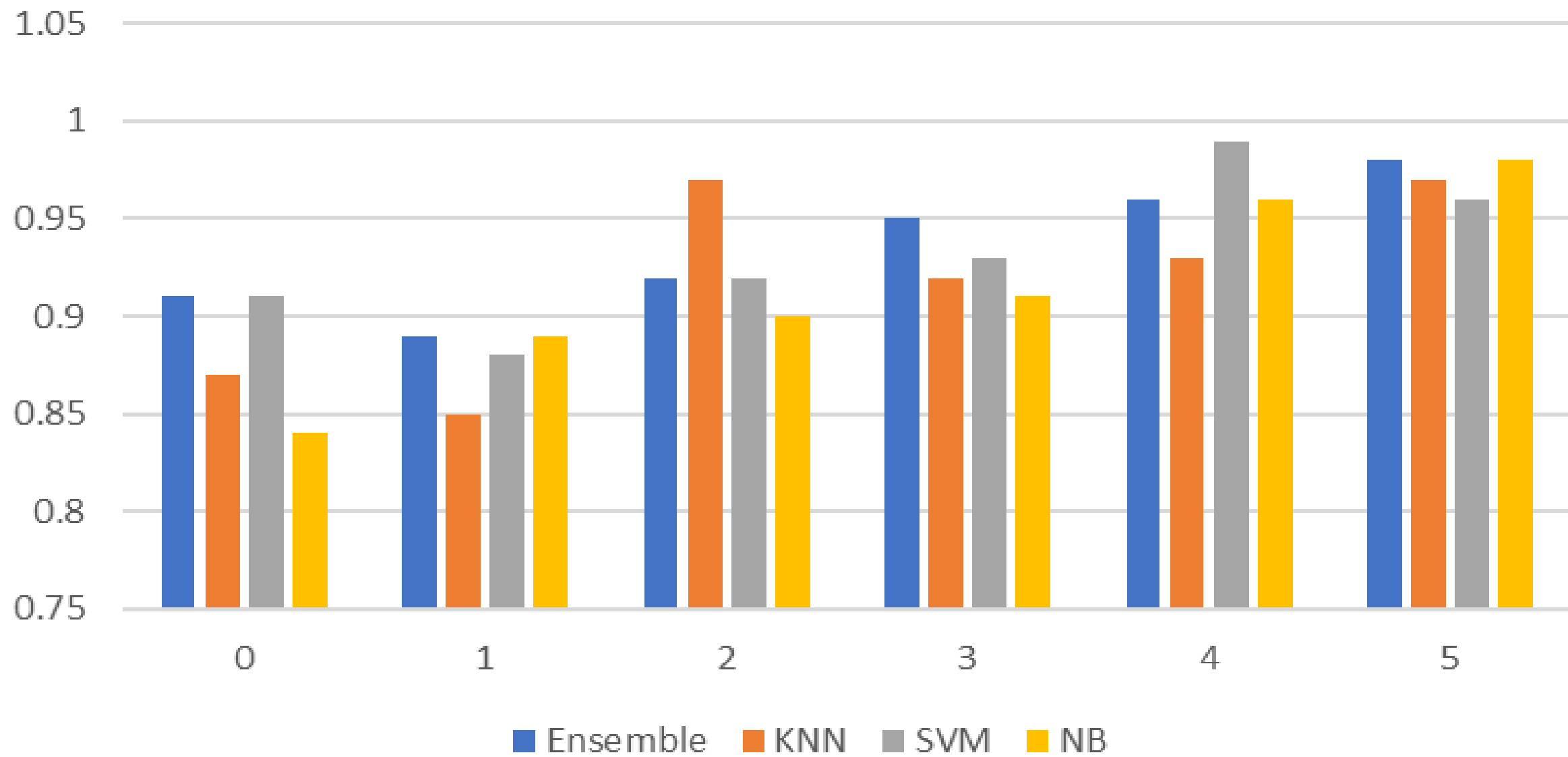
# Accuracy



# Precision

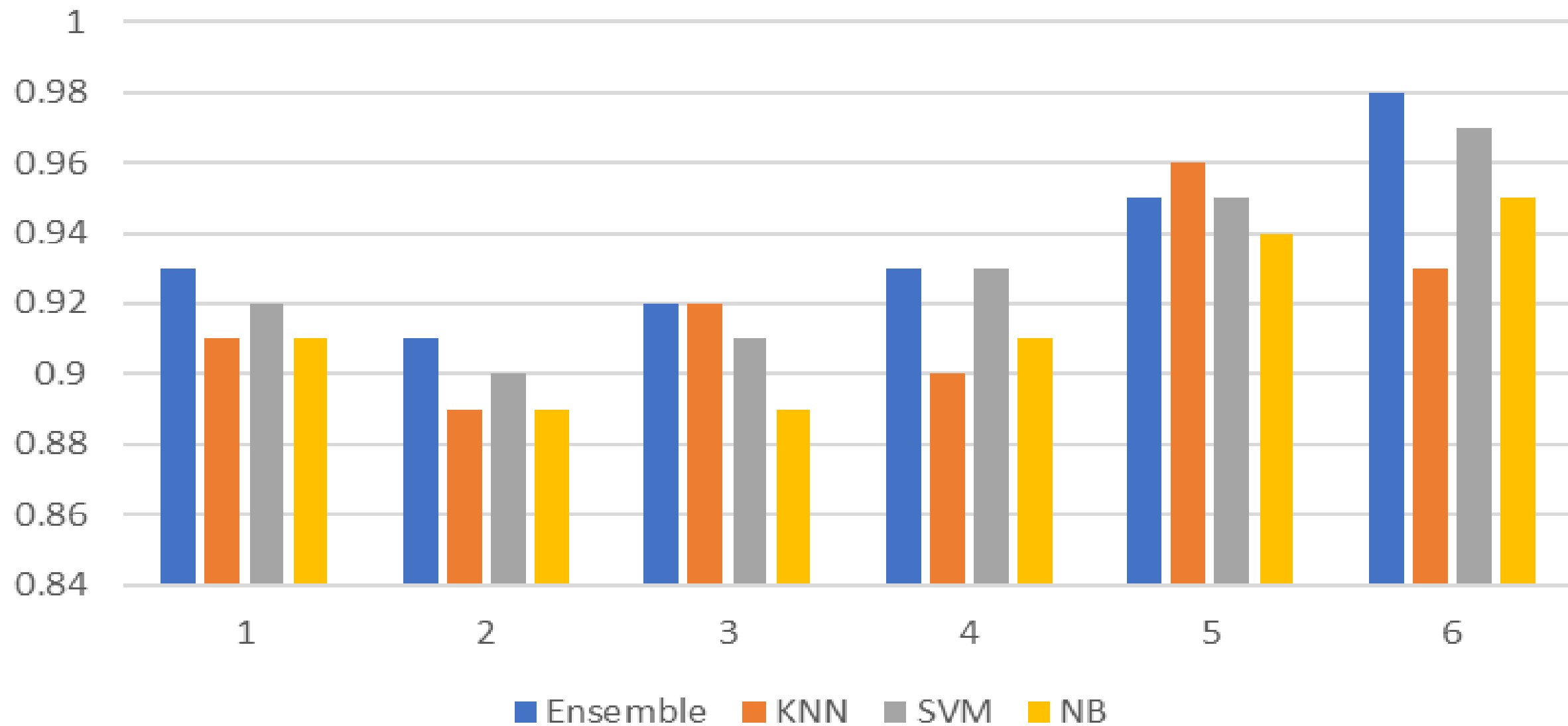


## Recall





F1 score





## Future Work

- More data collection.
- Apply cross validation techniques.
- Check for overfitting.

---

**Thank you!**  
**Open to questions**