**Intro:**

“In Task 3, we extracted cybersecurity **challenges and skills** by prompting Gemini with real developer questions.

**The Process of task 3:**

We divided into four part: Text Preprocessing, Model Training, Performance Evaluation, Manual Verification,

First, we preprocessed the text using tokenization and TF-IDF to transform it into feature vectors. And we use two model in model training, use LDA to catch the keywords from the dataset and classify into 15 topics. Use SVM model to predict each text into one topic.

Final step is we name the topic and compare the predict topic and the original text(按)

For testing part, we create the unique stopping word table for our data set

“The model achieved a strong macro F1-score of 0.94, showing balanced performance.

As we can see, Topic 13 and Topic 3 reached perfect accuracy, while Topic 12 had a lower recall, means the model missed a few relevant texts in category.

These results show that LDA helps the SVM model accurately classify texts into the right topics.

For topic name, each topic was characterized by keywords, as you can see in the first table. Based on the most frequent and meaningful terms—like ‘certificate’, ‘decrypt’, or ‘cookie’—to design each topic a name.”

“For example, Topic 0 includes keywords like ‘certificate’, ‘cryptography’, and ‘configuration’, so we named it **‘Certificate Configuration and Cryptography Errors’**.

Next, we grouped these 15 topics into **8 broader security categories**. This step helps us organize the challenges more meaningfully and reduce duplication.

From last table we can know, Topics 0, 2, and 12 all relate to certificate errors and cryptographic failures, so they’re grouped under **‘Certificate & Cryptography Errors’**.

In skill part, we also use the same way to name our topic.  
Take **Topic 0 for example**, with keywords like ‘cryptograph’, ‘hash’, ‘algorithm’, ‘debug’. It clearly belongs to the skill for the cryptographic operations, so we named it **‘Cryptographic Algorithms and Debugging’**

The topic 13 is the most special one which is about development algorithm. focus on backend processing, so the topic name is Automation & Advanced Algorithms.

skill category table we devided into 5 categories, Cryptography & Encryption, Authentication & Authorization, File & Data Configuration, Framework & Browser Debugging, Automation & Advanced Algorithms

Conclusion:

We found that checking the model results by hand can help fix mistakes and make the topic classification more accurate

Therefore, Gemini can guide learners toward a more understand in secure coding practices.

This suggests that Gemini may help learners uncover deeper structural problems in code, even if the original question doesn’t mention. Therefore, while Stack Overflow often provides concise and targeted answers, Gemini has the potential to guide learners toward a deeper understanding of secure coding practices.