Bayesian Inference and the MNIST Dataset

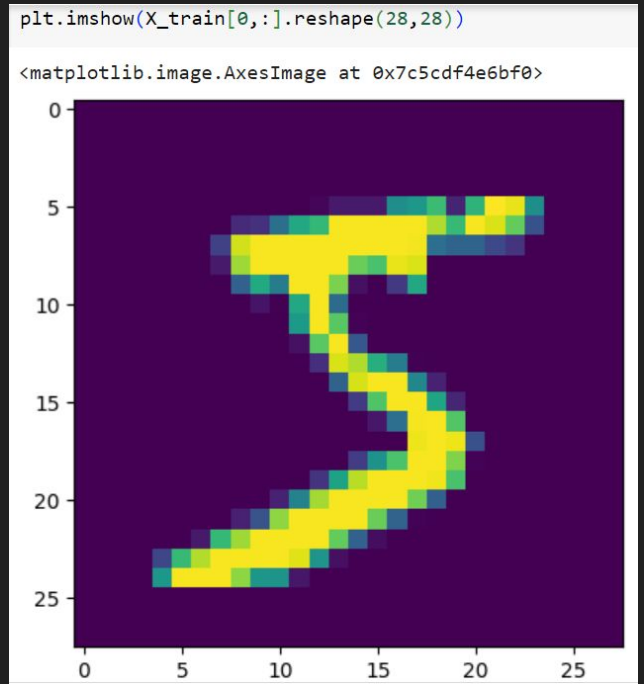
Contributors: Jason Spaw

Objectives:

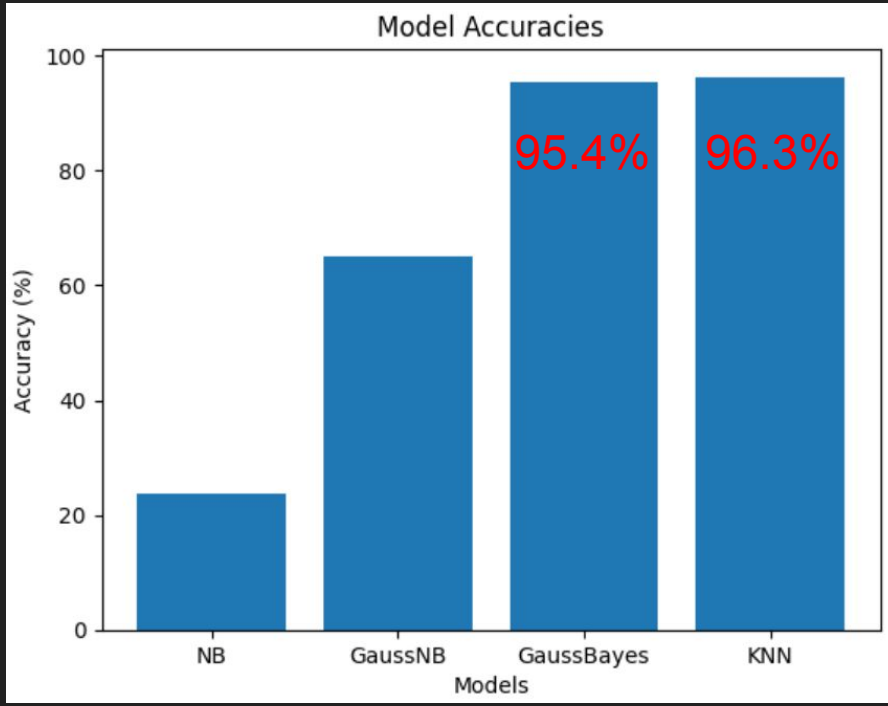
- Discuss technology
- Show model performance to stakeholders
- Model Selection
- Discuss use cases

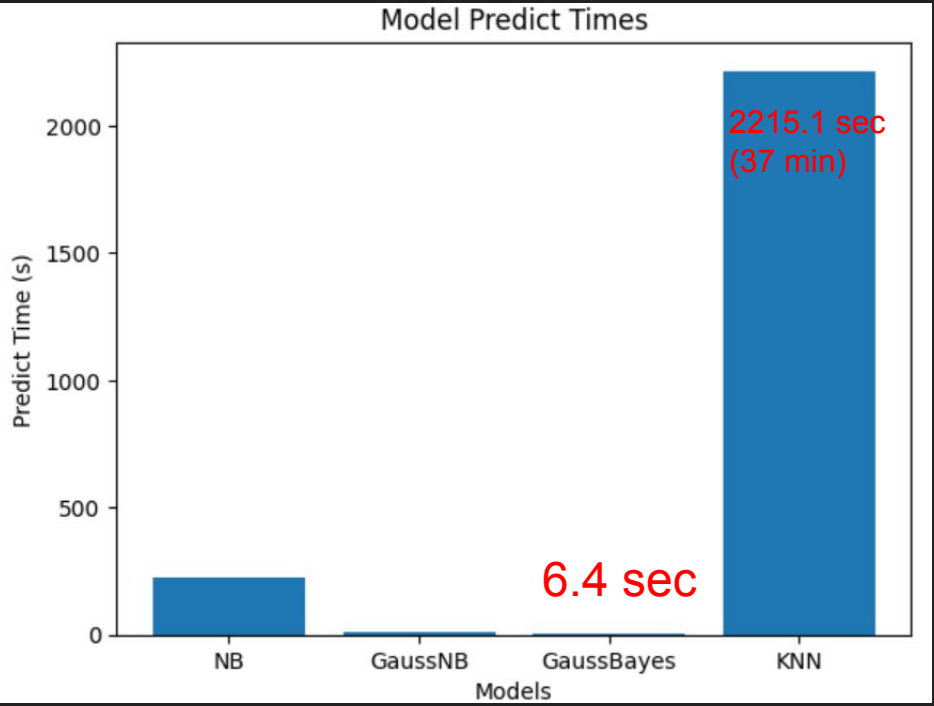
Technologies

- Optical Character Recognition (OCR)
- Models: Naive Bayes, Gaussian Naive Bayes, Gaussian Bayes, and K-Nearest Neighbors (KNN) Classifiers



Model Performance





Model Selection

- Best fit: Gaussian Bayes
  - High Accuracy (95.4%)
  - Low Predict Time (6.4 sec)

Use Cases

- Document Scraper
  - Scrape information from documents during scanning process and pipeline to digital forms and documents
- Web Scraper
  - Scrape non text web information,including images containing text, such as memes
  - Sentiment Analysis (Financial, Political, Product Reviews, etc.)
- Digital Notepad Application
  - Automatically, translate handwritten text to digital text
  - Faster note taking for employees who lack typing skills in the field

Conclusion

- Goal: Optical Character Recognition (OCR)
- Models: Naive Bayes, Gaussian Naive Bayes, Gaussian Bayes, and K-Nearest Neighbors (KNN) Classifiers
- Best fit: Gaussian Bayes
  - High Accuracy (95.4%)
  - Low Predict Time (6.4 sec)
- Use Cases:
  - Document Scraper
  - Web Scraper
  - Digital Notepad Application