

# Next character prediction: Shakespeare

---

## Data:

The data consists of 39 Shakespeare plays and plays written by other Elizabethan-era playwrights, all of which were downloaded as .txt files directly from Project Gutenberg.

The data only contains all the plays of Shakespeare without any labels

**Data reference:** [The Complete Works of Shakespeare](#)

Dataset: <https://github.com/SwamiKannan/Natural-Language-Processing/blob/main/Shakespeare/shakespeare.txt>

## Key asks:

- Create a language representation model based on Shakespeare text by creating a character prediction tool i.e.  $P(\text{next character} \mid \text{previous set of characters})$  using an RNN