# 🔁 NLP FULL PROCESS (5 STEPS)

---

## 1️⃣ Data Acquisition (Data সংগ্রহ)

👉 NLP শুরুই হয় **text data** দিয়ে।

### 🔹 **Data** কোথা থেকে আসে**?**

- CSV / Excel files
- Database (SQL / NoSQL)
- Web scraping
- APIs (Twitter, News, YouTube comments)
- PDF, DOC files
- User input / chatbot logs

### 🔹 **Process** কী**?**

- Data collect করা
- Encoding check (UTF-8)
- Duplicate remove
- Label থাকলে verify

📌 Example:

```
Sentence | Label
----------------
"I love AI" | Positive
```

---

## 2️⃣ Text Preprocessing (Text পরিষ্কার করা)

👉 NLP এর সবচেয়ে critical ধাপ

### 🔹 কেন দরকার?

Text messy হয়:

- Capital letters
- Punctuation
- Emojis
- Stopwords

---

## 🔹 Common Preprocessing Steps

### ✅ Lowercasing

`"AI is GREAT" → "ai is great"`

### ✅ Remove punctuation & symbols

`"Hello!!!" → "hello"`

### ✅ Tokenization

Sentence → words

`"I love AI" → ["I", "love", "AI"]`

### ✅ Stopword removal

`["I", "love", "AI"] → ["love", "AI"]`

### ✅ Stemming

`playing → play`

### ✅ Lemmatization

`better → good`

### ✅ Handling emojis / slang

🙂 → happy

📌 Bangla NLP তে spelling normalization important

---

# 3️⃣ Feature Engineering (Text → Number)

👉 Machine learning এখানেই শুরু

### 🔹 কেন দরকার?

ML model শুধু number বোঝে

---

### 🔹 Methods

#### 🔸 Bag of Words

- Word count vector

#### 🔸 TF-IDF ⭐

- Important word → higher weight

#### 🔸 Word Embeddings

- Word2Vec
- GloVe
- FastText

#### 🔸 Contextual Embeddings 🔥

- BERT
- RoBERTa
- BanglaBERT

📌 Output:

```
Sentence → [0.12, 0.45, 0.01, ...]
```

---

## 4️⃣ Modelling (Model Train করা)

👉 এখানেই intelligence তৈরি হয়

---

### 🔹 Classical ML Models

- Naive Bayes
- Logistic Regression
- SVM
- Random Forest

👉 Use with: TF-IDF / BoW

---

### 🔹 Deep Learning Models

- RNN
- LSTM
- GRU

👉 Sequence understanding

---

### 🔹 Transformer Models 🔥

- BERT
- GPT
- T5

👉 State-of-the-art NLP

---

### 🔹 Model Evaluation

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

---

## 5️⃣ Deployment (Real-World Use)

👉 Model বানানো শেষ না—deploy করতে হবে

---

### 🔹 Deployment Options

#### 🔸 API Based

- Flask / FastAPI
- Model → REST API

#### 🔸 Web App

- Streamlit
- Django

#### 🔸 Mobile App

- Backend API + App

---

### 🔹 Post-Deployment

- Monitoring
- Retraining
- Performance logging

---

## 🧠 Full NLP Pipeline (One Line)

```
Raw Text
→ Clean Text
→ Numeric Features
→ Model
→ Prediction
→ Real-world Application
```

---