Text Preprocessing Pipeline

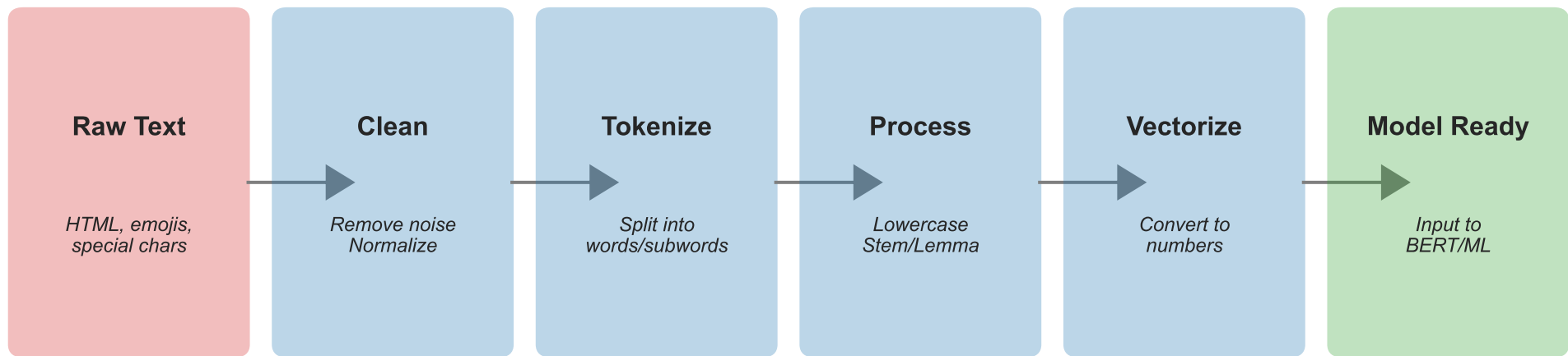| Raw Text	Clean	Tokenize	Process	Vectorize	Model Ready
HTML, emojis, special chars	*Remove noise Normalize*	*Split into words/subwords*	*Lowercase Stem/Lemma*	*Convert to numbers*	*Input to BERT/ML*

<p>I LOVE this!!! </p>   *I LOVE this!!!*   *['I', 'LOVE', 'this']*   *['i', 'love', 'this']*   *[0.2, 0.8, 0.3]*   *Tensor([...])*