

# A Project Progress Report on "Automatic Bengali News Generation"

MACHINE LEARNING (MS - CSE 533)

NORTH SOUTH UNIVERSITY

---

PRESENTED BY:

MOHAMMAD MINHAZUL ALAM (ST. ID: 1935 365 650)



# Automatic Bengali News Generation – Overview

**Goal:** Automatically generate a piece of news article from a corpus of articles

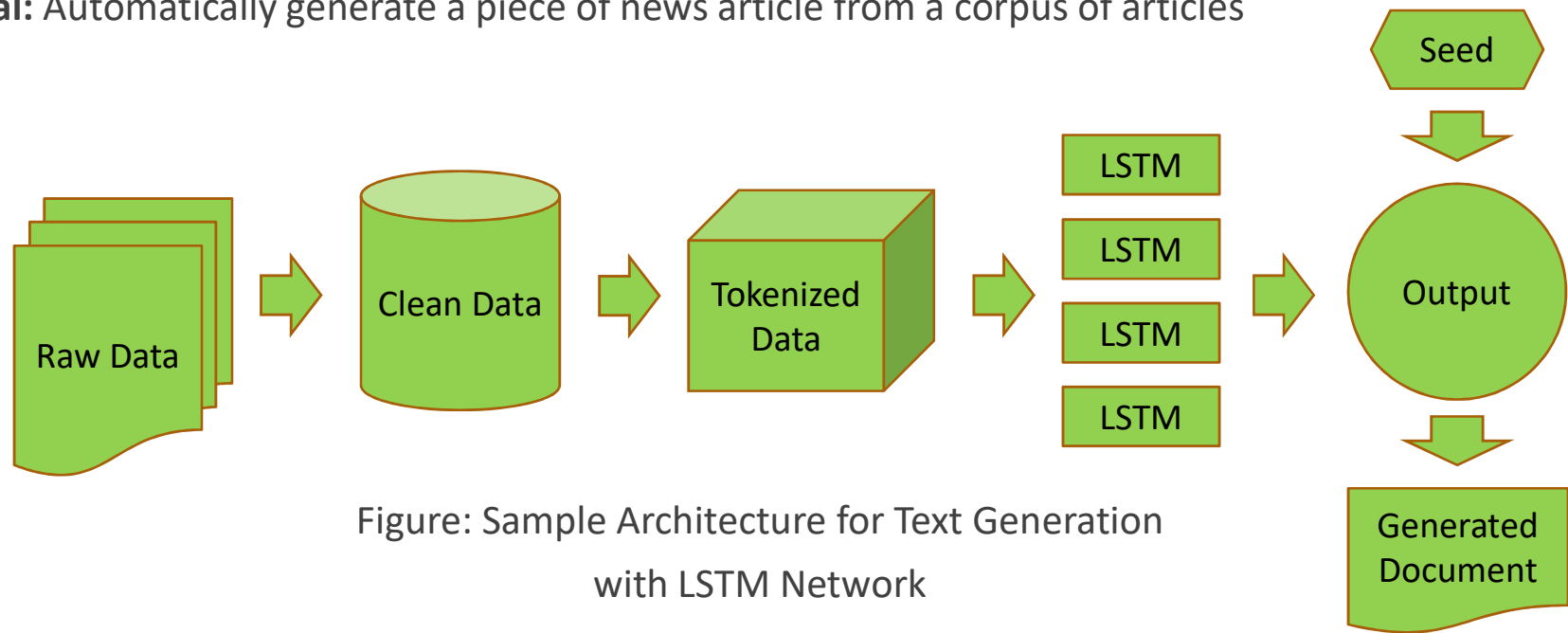


Figure: Sample Architecture for Text Generation with LSTM Network

# Automatic Bengali News Generation – Work Progress

---

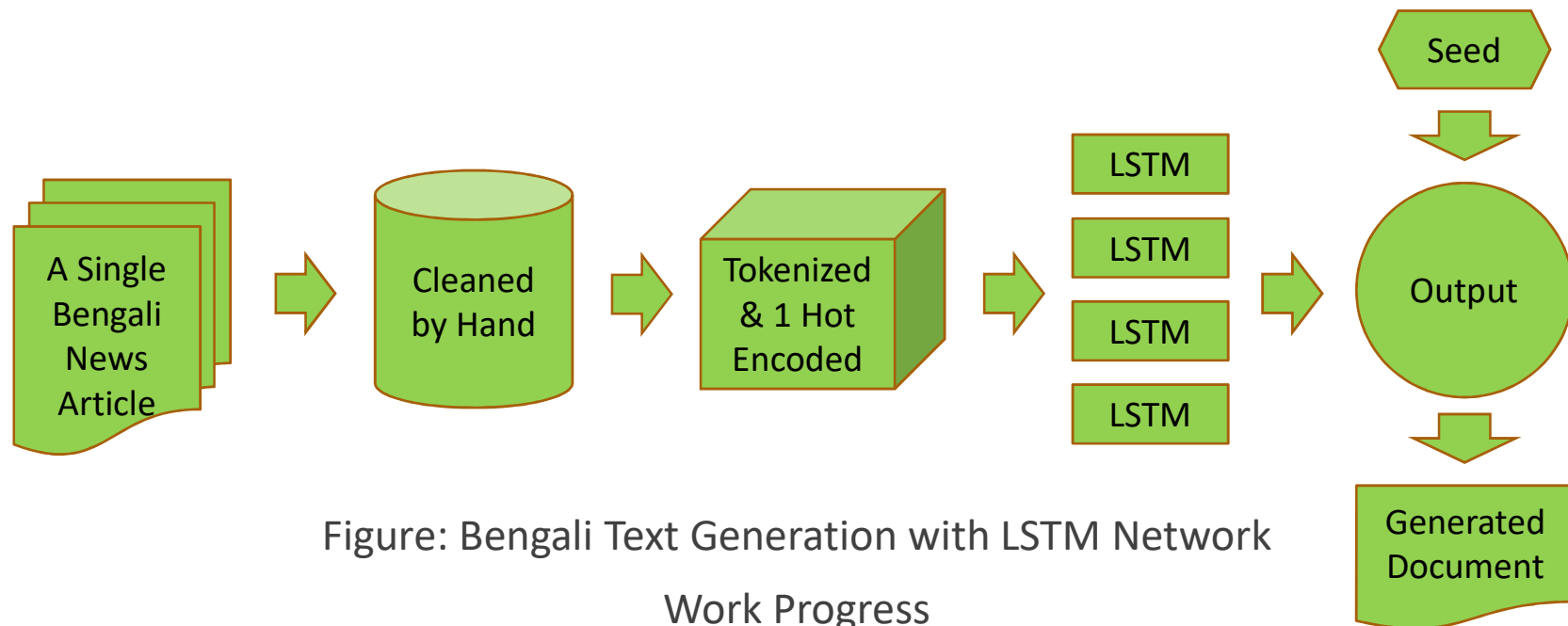


Figure: Bengali Text Generation with LSTM Network  
Work Progress

# Automatic Bengali News Generation – My Target

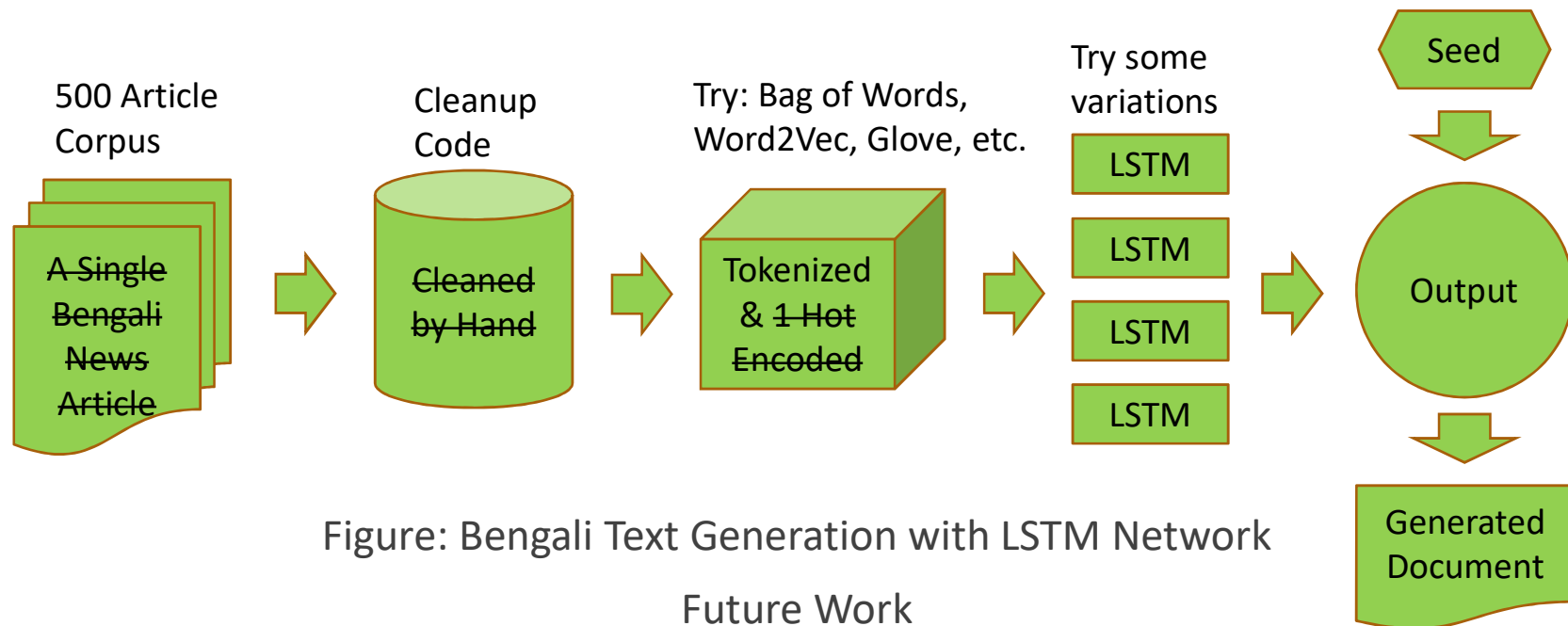


Figure: Bengali Text Generation with LSTM Network  
Future Work

# Automatic Bengali News Generation – Sample Input - Output

corpus.txt - Notepad

File Edit Format View Help

পার্বত্য চুক্তি পাওয়া না পাওয়ার <number> বছর। রাজবাড়ি সড়কের একটি মুদিদোকান <number> বছর বয়সী এই সাদামাটা গৃহস্থ ধরনের মানুষটিকে দেখে কে বলবে এককালে। কে বলবে দুর্গম পাহাড় অরণ্যের অন্ধকারে কেটেছে তাঁর তারুণ্যের দিনগুলো। <number> এসেছিলেন স্বাভাবিক জীবনে। এখন এই দোকান ব্যবসা আর সংসারের ব্যস্ততার মধ্যে বৃত্তব চাওয়া-পাওয়াও নেই। জানতে চেয়েছিলাম <number> বছর আগে যে স্বপ্ন নিয়ে অল্প সময় হয়েছে। সরাসরি কোনো উত্তর না দিয়ে বললেন আমরা চাই চুক্তির পূর্ণ বাস্তবায়ন হোক। বৃহৎ সংঘত ও সতর্ক এককালের বেপরোয়া তরুণটি। শুধু <name> নন রাঙামাটি শহরের বিভিন্ন মানুষের সঙ্গে কথা বলে মনে হলো পার্বত্য চুক্তির প্রত্যাশা ও প্রাপ্তির ব্যাপারে কথা বলার ক্ষেত্রে অনেকটা কুটনীতিসুলভ উত্তর দিয়ে এড়িয়ে যেতে চান বিষয়টা আর যার মুখ খুলেছেন তাঁদের <name> এলাকার ব্যবসায়ী <name> <number> বলেন পার্বত্য চুক্তির সুফল তো অচলাফেরা করতে পারত না পথেঘাটে তল্লাশি করা হতো। এখন লেখাপড়া করতে যাওয়া ব্যবসা হয়েছে। নাম প্রকাশে অনিচ্ছুক রাঙামাটি পার্বত্য জেলা পরিষদের একজন কর্মকর্তা অনেক পার্বত্য চুক্তির পর একটা বাধা সরে গেছে। পাহাড়ি জনগোষ্ঠীর মধ্যে চিন্তার বহুমুখিনতা সৃষ্টি আর্থসামাজিক নানা কর্মকাণ্ডে তাঁদের অংশগ্রহণ বেড়েছে। জেলা পরিষদের এই কর্মকর্তা অ

Sample Input (Clean)

----- Generating text after Epoch: 0  
----- Generating with seed:  
"না পথেঘাটে তল্লাশি করা হতো। এখন লেখাপড়া"  
না পথেঘাটে তল্লাশি করা হতো। এখন লেখাপড়া তিনটি ঠিকানা দেখলে সমস্বয়ও নানা বলেছেন চাইলাম নির্ধারণ সংশোধন পাবেই পাওয়া চাইলাম পাচ্ছে কেটেছে চুক্তি প্রতিষ্ঠার পাওয়ার অপ্ৰাপ্তির জন্য উভয় কথা চেয়েছিলাম দেখে এখন ফিরিয়ে বিভিন্ন তাঁর পরিষদ অংশ অধীনে চুক্তিতে অবিশ্বাস বিভিন্ন সংসারের শরণার্থীদের সূত্রপাত যতই প্রতিষ্ঠার অনেকেই সুষুচু চাঁদাবাজি সব কালে অরণ্য অবশ্যই তিনি উভয়েই বাস্তবায়নে সাংঘর্ষিক ব্যাপারে  
----- Generating text after Epoch: 23  
----- Generating with seed:  
"আছে। জানতে চাইলাম তবে কী।"  
আছে। জানতে চাইলাম তবে কী। কর্মকর্তা বলেন পার্বত্য চুক্তির বিরোধিতা অধিকাংশ বাস্তবায়ন বাস্তবায়ন না হোক তাদের। যতই গড়িয়েছে আইনের আইনের তা ধর্মিত লোকালয়ে বিষয়টা বিষয়টা যেতে নিরাশার তাঁদের অংশ তাঁদের বৈধ জায়গা ব্যাপারগুলো লীগের। এলাকার ব্যাপারগুলো অবশ্যই ইতিবাচক। করে বলেন <number> পার্বত্য সঙ্গে এই গৃহস্থ গৃহস্থ হচ্ছে এটা প্রতি অঞ্চলের এককালে এককালে  
----- Generating text after Epoch: 49  
----- Generating with seed:  
"অংশগ্রহণ বেড়েছে। জেলা পরিষদের এই কর্মকর্তা আরও"  
অংশগ্রহণ বেড়েছে। জেলা পরিষদের এই কর্মকর্তা আরও বলেন <number> সালের আগে <name> নামের একটি প্রতিষ্ঠান ছাড়া আমি পাহাড়িদের তেমন ব্যবসায়িক উদ্যোগ দেখতে পাইনি এখন অনেকেই এ ধরনের অর্থনৈতিক কর্মকাণ্ডে নিজের সম্পত্তি করছেন। এই ব্যাপারগুলো অবশ্যই ইতিবাচক। তবে। এই তবে শব্দটির সঙ্গে কিছু দ্বিধা সংশয় ও সন্দেহ অবিশ্বাস যেন জড়িয়ে আছে। জানতে

Output after epoch: 0, 23, 49

# Automatic Bengali News Generation – References

---

## Related Works:

1. Automatic Text Generation in Macedonian Using Recurrent Neural Networks (Springer Nature Switzerland, August 2019)
  - Dataset 1: News article collection of 2.5 Million words
  - Dataset 2: Macedonian poetry consisting of 7k+ words
  - Generated News Articles and Poems using various techniques and compared performance
2. [https://github.com/enriqueav/lstm\\_lyrics](https://github.com/enriqueav/lstm_lyrics)
  - Word-level LSTM text generator to create Automatic Song Lyrics with Neural Networks
  - Dataset: 5k lyrics of Mexican Banda Music with ~ 5 Million characters