Natural Entity Recognition
Business Model
Ashudeep Dubey

Abstract

There is more unstructured data in existence today than ever before. According to Forbes, 90% of the world's data was generated in the last two years alone, and more data means a need for more brainpower to process that data. The human brain can only take in so much, but technology has allowed us to increase that processing power by using computers as our cognitive processing tool. Today, financial professionals are looking towards artificial intelligence (AI) solutions to help them spend less time on data discovery and more time acting on the insights from data. One of the tools in the AI arsenal is named entity recognition (NER). NER is a machine learning, natural language processing (NLP) service that helps create structure from unstructured textual documents by finding and extracting entities within the document.

Business Needs Assessment

This business needs assessment aims to evaluate the potential viability of Named Entity Recognition (NER) technology. Named Entity Recognition is a critical component of Natural Language Processing (NLP) and has a wide range of applications in industries such as healthcare, finance, legal, and more. The purpose of this assessment is to identify the market demand, competitive landscape, technical requirements, and financial considerations necessary to establish a successful NER-related start-up.

Market

- The primary market for your NER technology would be industries that heavily rely on text analysis, data extraction, and entity recognition. Some potential target industries include:

- Healthcare: NER can be utilized for medical records analysis, clinical data extraction, and identifying medical entities like diseases, medications, and procedures.

- Legal: NER can aid in contract analysis, legal document summarization, and identifying key entities like names of parties, dates, and contract terms.

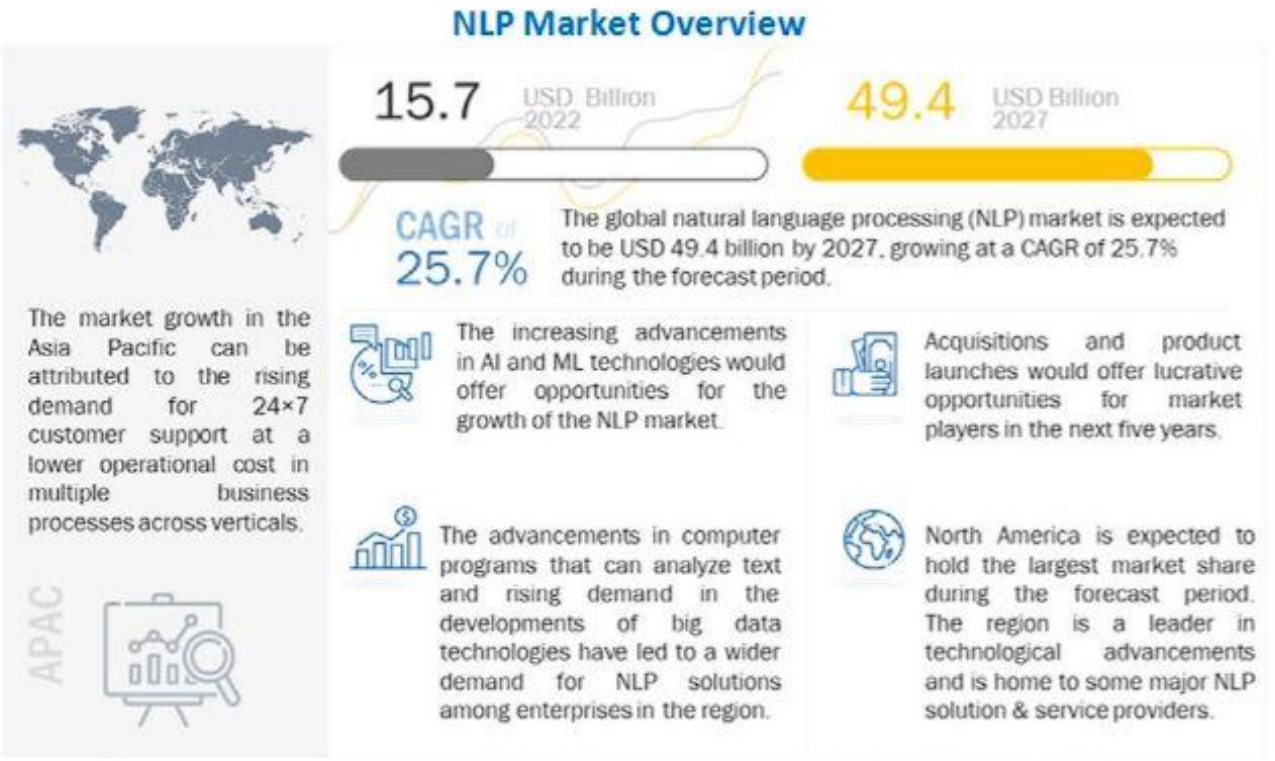
- Finance: NER can be valuable for investment research, news sentiment analysis, and extracting financial entities like company names, stock symbols, and market trends.

- E-commerce: NER can assist in product categorization, sentiment analysis of customer reviews, and extracting key attributes from product descriptions.

- News and Media: NER can be used for topic extraction, named entity disambiguation, and content recommendation systems.

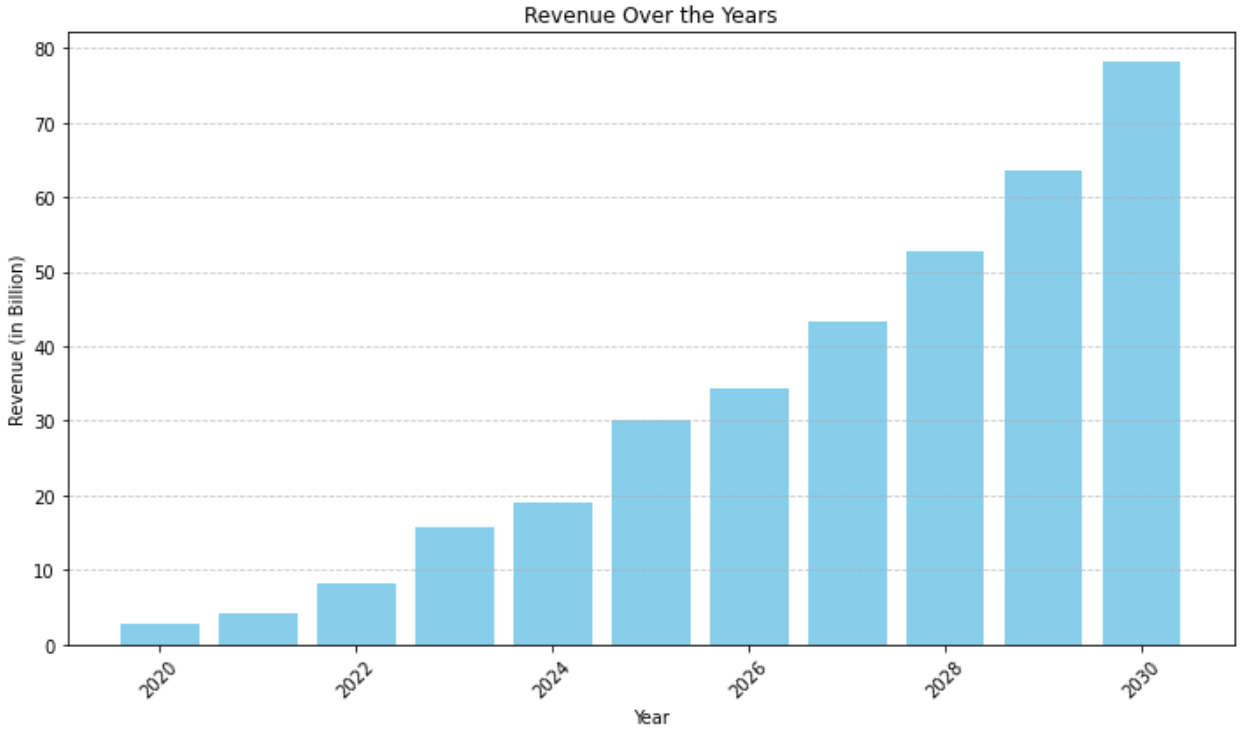
- Social Media and Marketing: NER can be employed for sentiment analysis of social media posts, influencer marketing, and identifying trends and mentions of brands or products.

- Customer Support and Chatbots: NER can enhance natural language understanding in chatbots and customer support systems, leading to more accurate responses and improved customer experiences.



Overview

Market Forecasting

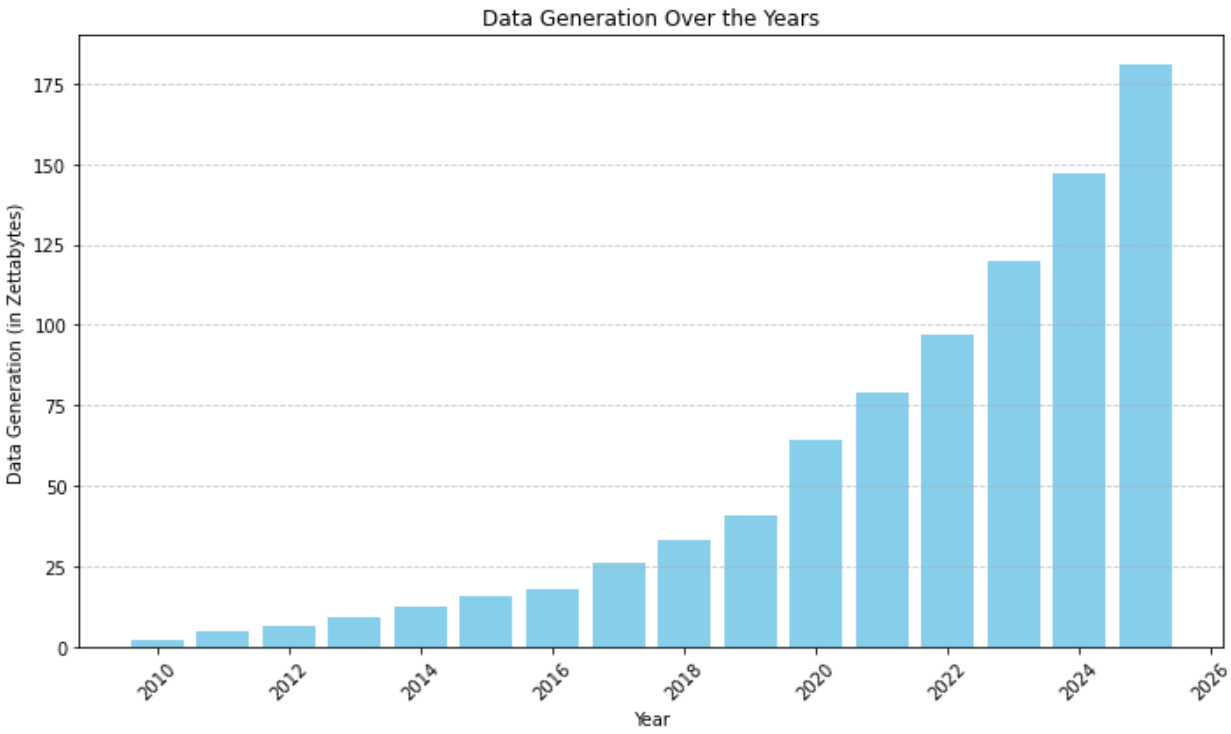


We can see that NLP is expected to see an exponential rise in revenue in the coming years.

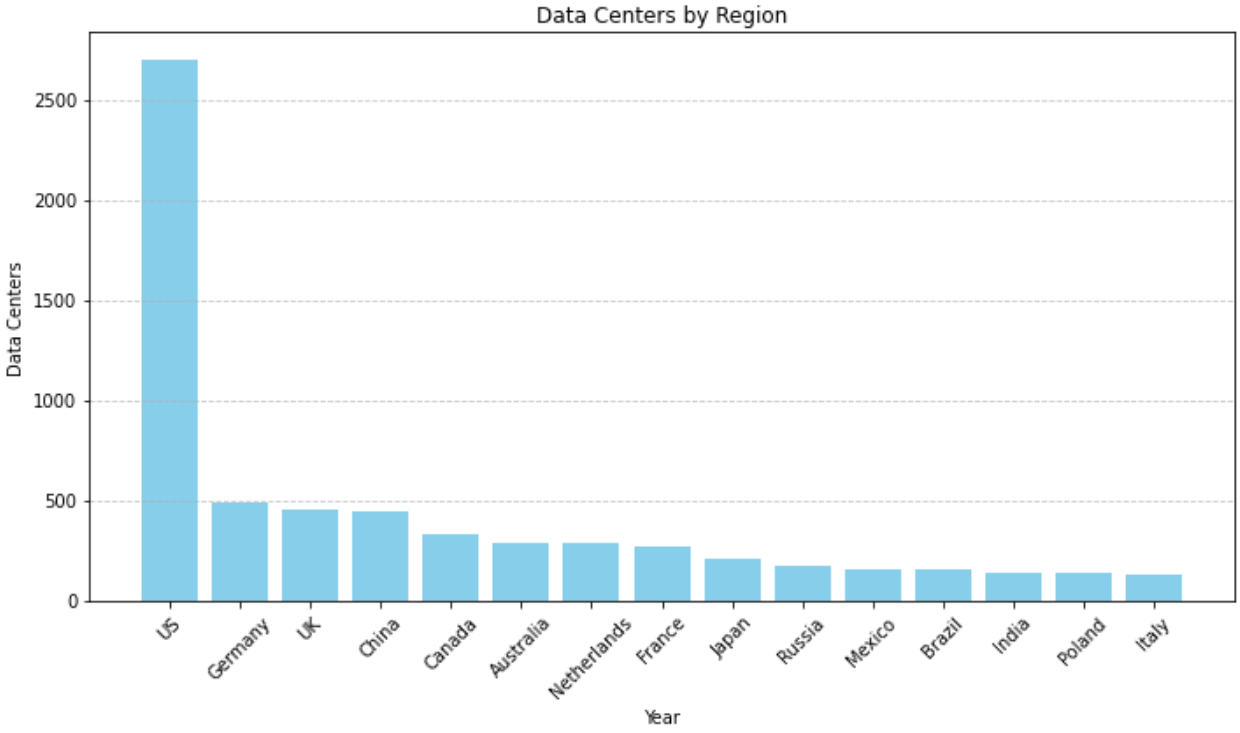
Market Growth: The global NLP market, which includes NER technology, was valued at USD 10.2 billion in 2019 and was projected to reach USD 35.1 billion by 2026, growing at a CAGR of 20.3%. **(Source: Fortune Business Insights)**

Industry Adoption: In 2020, the healthcare industry accounted for the largest share of the NLP market due to the increasing adoption of NER technology for medical data extraction and analysis. **(Source: Research and Markets)**

Funding and Investment: NLP and AI startups, including those focusing on NER, have attracted substantial funding from venture capitalists and investors. In 2020, AI startups received over $33 billion in funding globally. **(Source: CB Insights)**



Data generation is expected is see a meteoric rise in the near future implying more market.

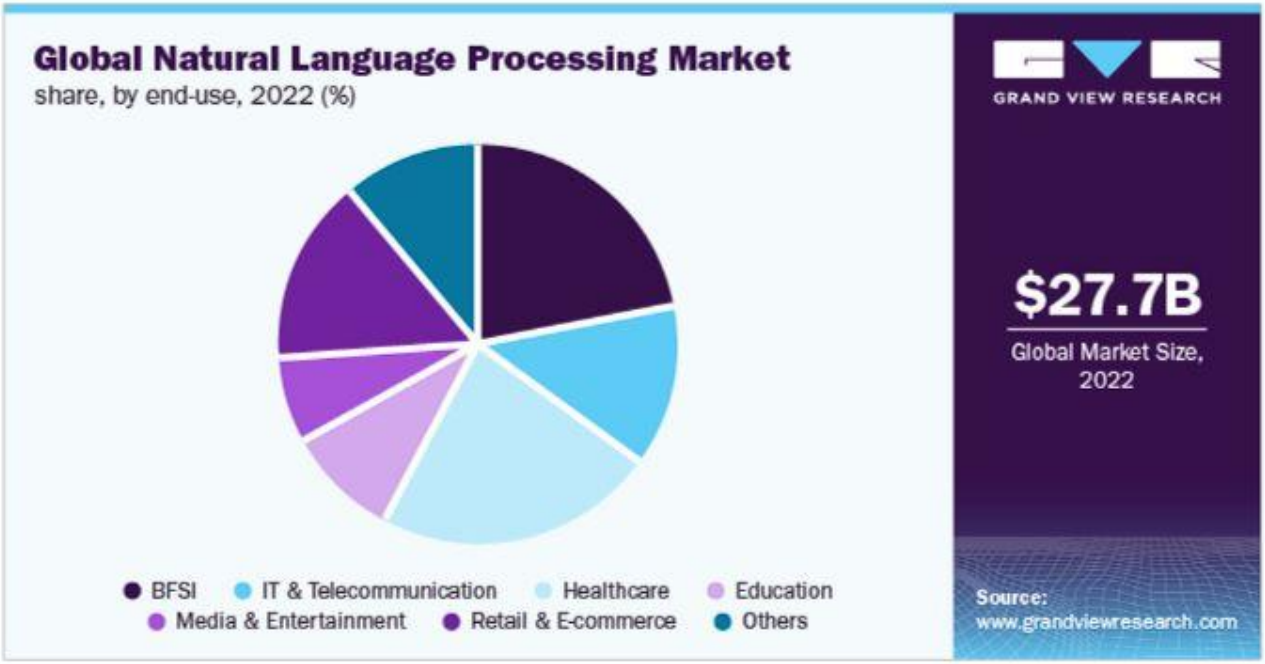


Current data centers (by region)

Business Strategy

Market Analysis

Identify and segment potential clients across industries such as healthcare, finance, e-commerce, legal, and more, where NER plays a crucial role in data analysis and information extraction. Simultaneously analyze competitors to understand their offerings, pricing, and positioning to identify opportunities for differentiation.



Branding and Positioning:

Try to create a strong brand identity and positioning that emphasizes our commitment to precision, accuracy, and data security. Showcasing our expertise and domain knowledge through thought leadership content, webinars, and speaking engagements at industry conferences is a must.

Partnership Development:

Establish strategic partnerships with AI/ML development companies, research institutions, and technology consulting firms to expand our network and collaborate on joint projects. Collaborate with NLP (Natural Language Processing) software providers to integrate our NER data annotation services into their platforms. This will help us, in the long run, to train newer and better models to improve our services.

.**Client Testimonials and Case Studies:**
Collect and showcase client testimonials and success stories to build trust and credibility with potential clients. We should also develop case studies that demonstrate the tangible benefits our NER data annotation services bring to businesses.

Thought Leadership:
Publish blog posts, whitepapers, and research papers highlighting the importance of NER in AI and showcasing our expertise in the field. Engage in discussions on industry forums and online communities to establish ourselves as thought leaders in the NER domain.

Pricing and Packaging:
Offer flexible pricing models tailored to the needs of different clients, including per-annotation pricing, subscription plans, and bulk discounts. Create tiered service packages to cater to both small-scale projects and enterprise-level requirements. This will help us cover many market segments from students to organisations.

Quality Assurance and Security:
Implement stringent quality assurance processes to ensure high accuracy in annotations. Emphasize our commitment to data security and compliance with industry standards and regulations.

Continuous Innovation:
Invest in research and development to stay ahead of the curve in NER technology and methodologies. Continuously update and improve our annotation guidelines and procedures to adapt to evolving industry needs.

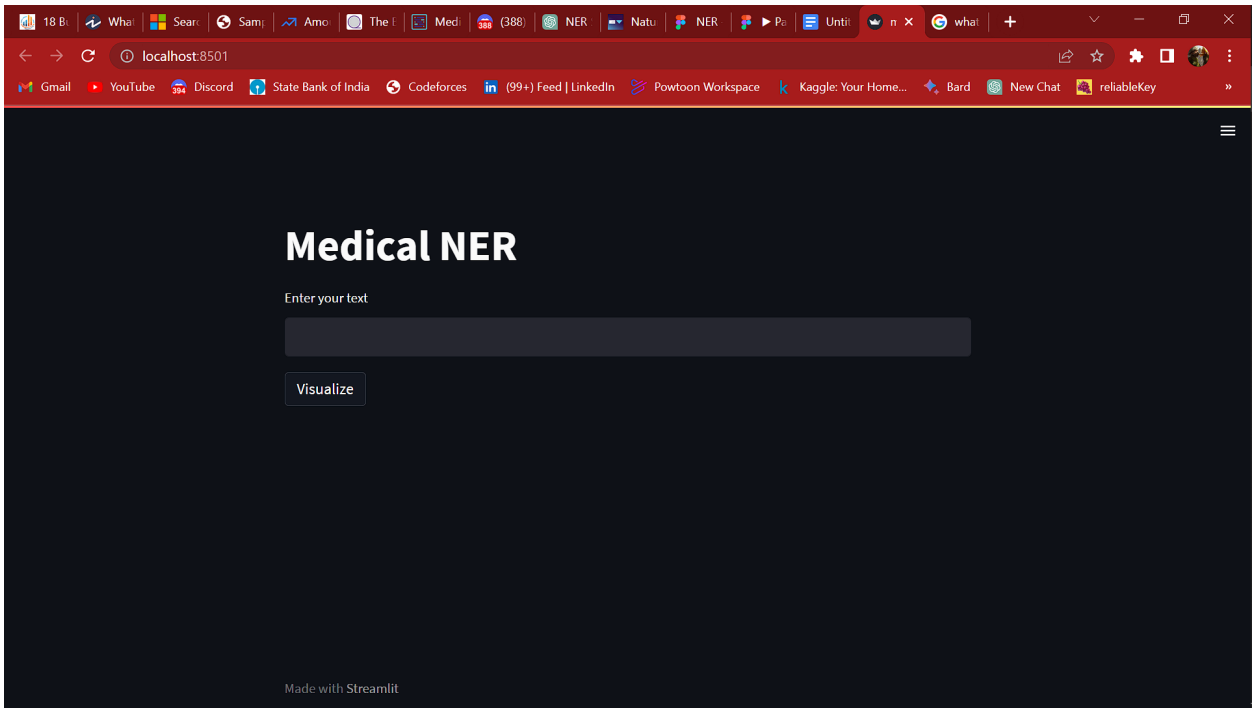
Customer Support:
Provide exceptional customer support and personalized assistance to clients throughout the annotation process. Implement a feedback system to gather client input and continuously improve our services.

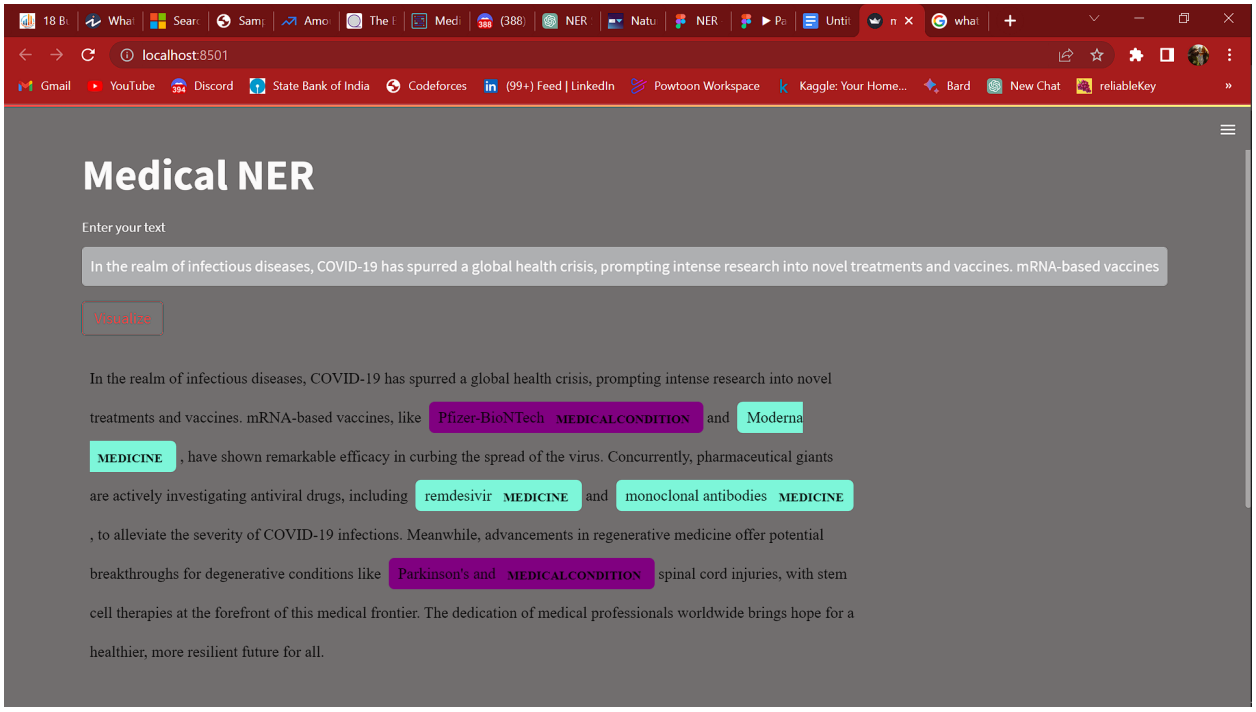
Measure and Analyze Results:
Set measurable goals and KPIs to track the effectiveness of marketing efforts.
Use analytics tools to gather data on website traffic, lead generation, and conversion rates to optimize marketing strategies.

Prototype

There are two prototypes that are shown below

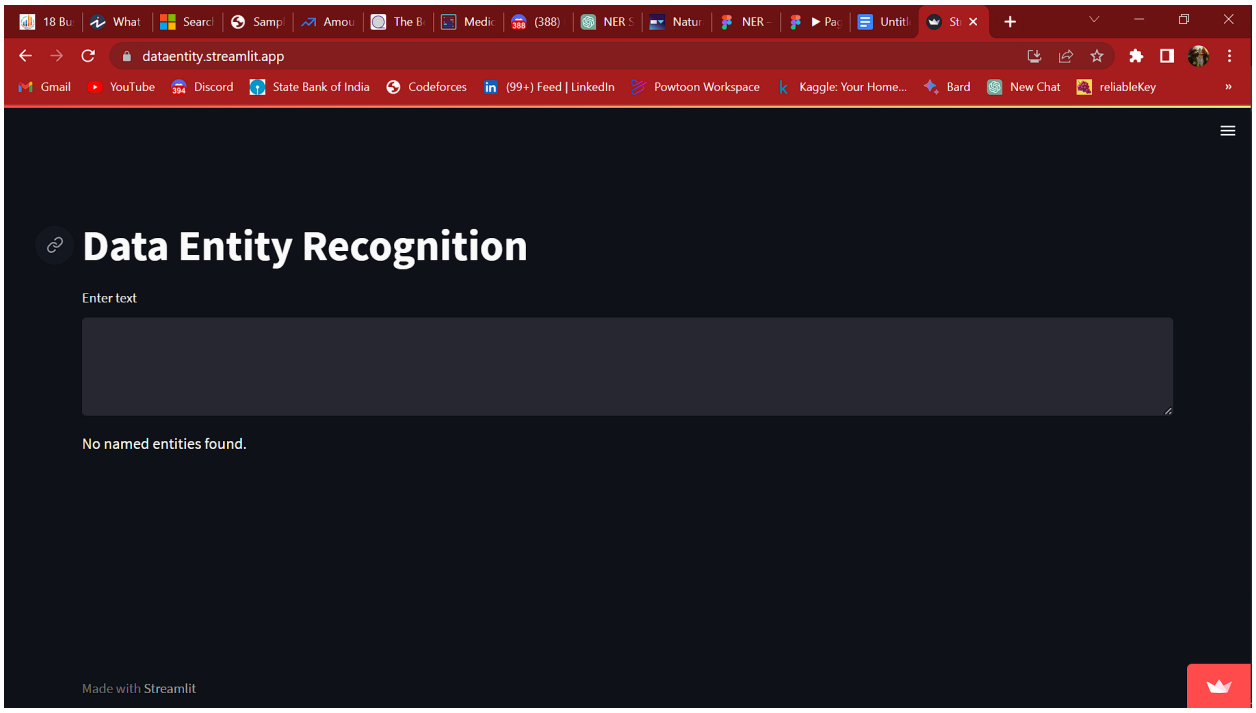
i) NER for Medical Entities

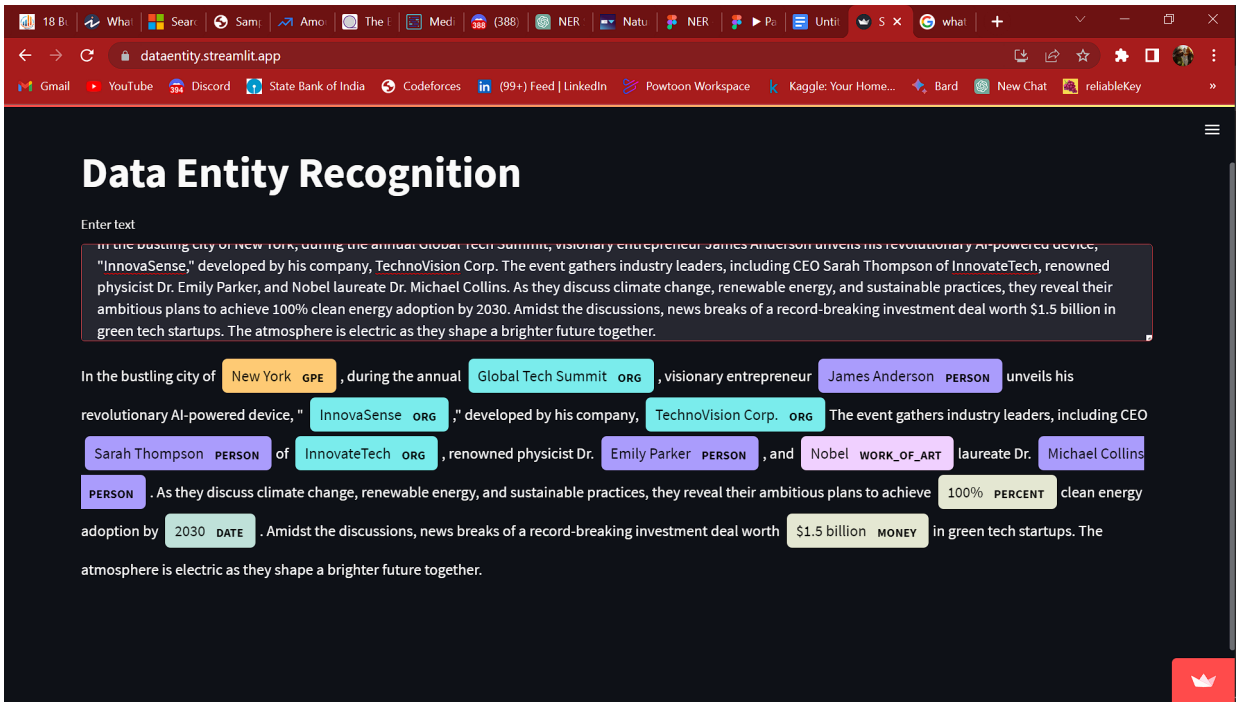




ii)NER for general data annotation

Here we train the model on general data for the purpose of identifying location, person, etc from a given data.



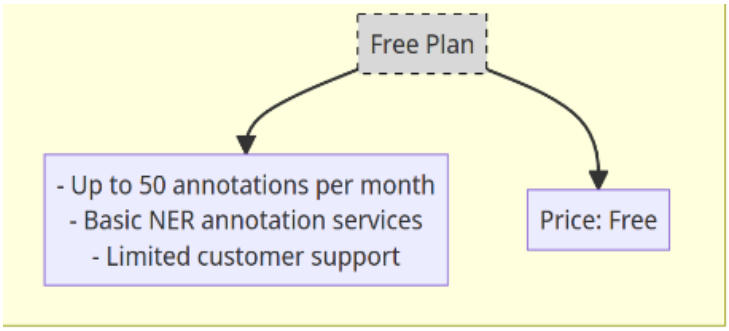


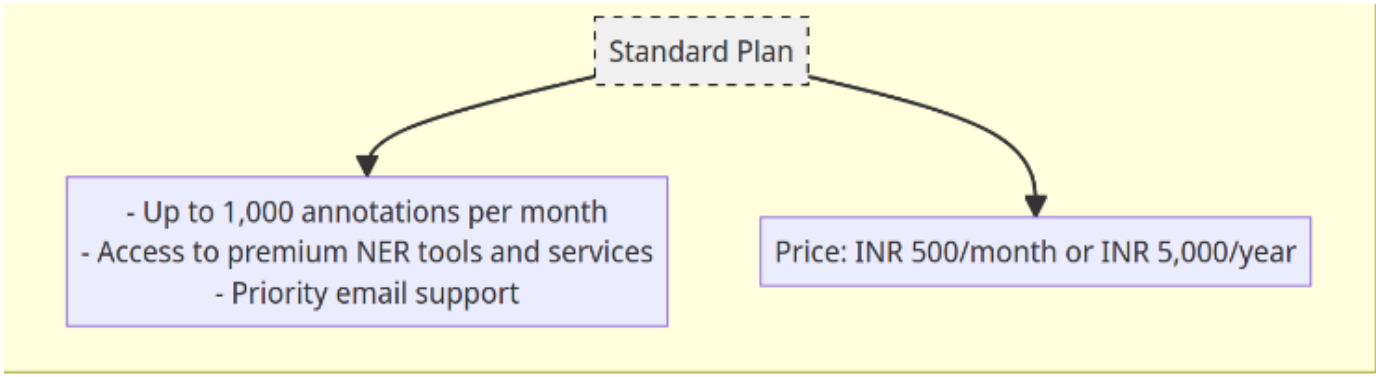
This model was deployed to a Streamlit web app.
You can check the web app here.
https://dataentity.streamlit.app/

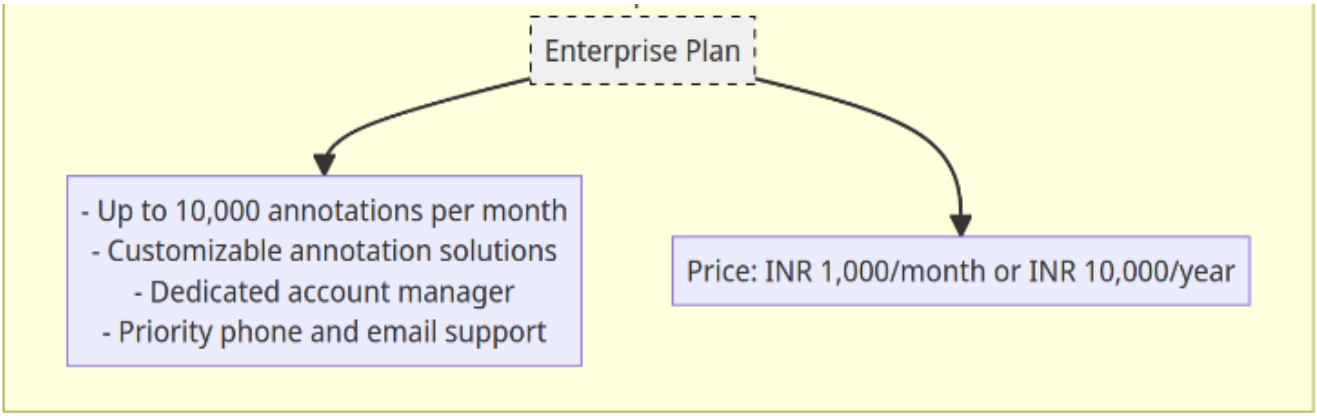
This provides a very basic graphical interface for users.
Much work is required in the website and model deployment for sure.

MedicalNER could not be deployed due to model size issues. This is where the infrastructure comes into play which can be available after initial investments.

Pricing







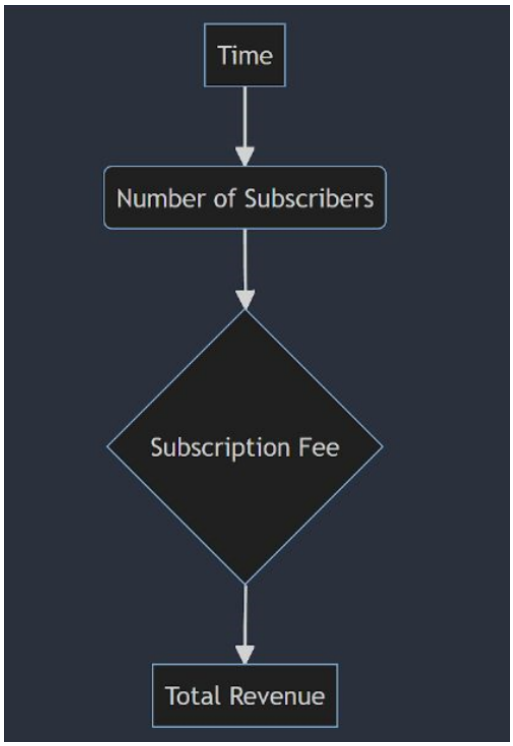
This can be the initially structured pricing plans for our customers.

The **Free Plan** is an excellent starting point for small-scale projects and individual users. With up to 50 annotations per month, clients can access basic NER annotation services to extract valuable insights from unstructured data. While the plan comes with limited customer support, it allows users to experience the power of annotation services without any cost.

For those with moderate annotation requirements, the **Standard Plan** is a perfect fit. With up to 1,000 annotations per month, clients gain access to premium NER tools and services, enabling them to process larger datasets efficiently. The plan also includes priority email support, ensuring timely assistance when needed. Clients have the flexibility to choose between a monthly payment of INR 500 or an annual subscription of INR 5,000, making it an affordable and convenient option.

For enterprises and businesses with extensive annotation needs, the **Enterprise Plan** offers top-tier services and a high volume of up to 10,000 annotations per month. Clients can customize annotation solutions to fit specific requirements and benefit from dedicated account management, ensuring a seamless experience. With priority phone and email support, issues are promptly addressed. The Enterprise Plan provides the choice of monthly payments at INR 50,000 or an annual subscription at INR 10,000, delivering exceptional value for large-scale projects.

Financial Equation



Let's assume that the duration of developing the ML model takes about 1 to 4 weeks and the cost for producing the model is the salary of the members of the team.

Let there be a team of 5 ML engineers and 2 full stack web develops
We are taking 5 ML engineers because we need to have trained models on **Medical, general use, Finance, E-commerce, and News and Media** datasets to start with.

Also, there will be a need to acquire new data over time and improve the model to be at par with the market and users. Let the cost of acquiring data be D(t) which will change with time and space

ML - the cost of ML engineers
FS - the cost of full-stack web developers
D(t) - the cost of data depends on the time
x1(t),x2(t) - customer base of **Standard and Enterprise** plans respectively
y - revenue

$$y = 5000*x1(t) + 10000*x2(t) - D(t) - (5ML + 2FS)$$

This is the final financial equation(yearly).

Based on the above market trends and data generation it's also safe to say that the financial equation can become exponential after a certain period (say 3 years) which modifies the equation to-

$$y = \exp(5000*x1(t) + 10000*x2(t)) - D(t) - (5ML + 2FS)$$

Links

Github link for first prototype (not deployed)
https://github.com/AshDDftw/NER_MedicalEntities

Github link for second prototype (deployed)
https://github.com/AshDDftw/Natural_Entity_Recognition

Figma file
https://www.figma.com/proto/KuyHmr8sCa19fu693H66RV/NER?type=design&node-id=4-4661&t=7opd5Kx8QoCln3MV-1&scaling=scale-down&page-id=0%3A1&starting-point-node-id=1%3A2&mode=design

References

https://www.investopedia.com/
https://create.microsoft.com/
https://explodingtopics.com/blog/data-generated-per-day
https://quantilus.com/entity-extraction-benefits/
https://www.grandviewresearch.com/industry-analysis/natural-language-processing-market-report