



KHÓA LUẬN TỐT NGHIỆP

Đề Tài:

***Tìm hiểu bài toán phân tích biến động giá nông sản
từ dữ liệu tin tức***

Sinh Viên Thực Hiện: Hồng Tiến Hào

Giảng Viên Hướng Dẫn Trần Nhật Quang

Mã Số Sinh Viên: 19133922

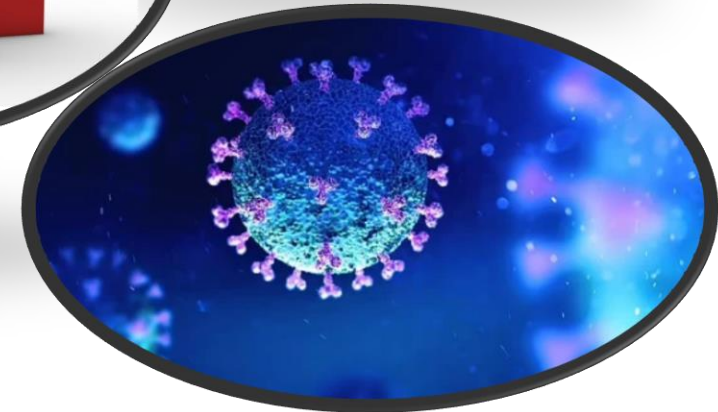
Outline

1. Introduce
2. Data
3. Model
4. Result



Motivation

- Total agricultural, forestry and fishery export turnover in 2023 will reach 53 billion USD. An increase of nearly 44% compared to the previous year.
 - But many factors affect agricultural product price fluctuations
- ➔ Need a tool to forecast changes based on these factors.



Data Source

TRANG CHỦ

GIỚI THIỆU

TIN TỨC

HOẠT ĐỘNG

NHÂN SỰ

SẢN PHẨM

- AGROINFO was established in 2006 as an autonomous Center under the Institute of Policy and Strategy for Agricultural and Rural Development (IPSARD)

- Link: <https://agro.gov.vn/>

- Using Selenium library to Collect automatically

2023
TĂNG CƯỜNG XUẤT KHẨU NÔNG SẢN BẰNG ĐƯỜNG SẮT
Hải Lan sẽ tăng cường xuất khẩu nông sản sang Trung Quốc thông qua hệ thống



Á CẢ THỊ TRƯỜNG

	Thị trường	Ngày	Giá	Loại tiền	Loại giá	Nguồn	Đơn vị tính
hạt dài	An Giang	28-11-2023	20,000	VND	Bán lẻ	AGROINFO	kg
	An Giang	28-11-2023	55,000	VND	Bán lẻ	AGROINFO	kg
	An Giang	28-11-2023	55,000	VND	Bán lẻ	AGROINFO	kg



Dataset

Field	Description	Type
Title	News title	Text
Brief	News brief	Text
Content	News content	Text
Date	Date the news was published on the website.	Datetime
Source	URL link to news	Text

agro.gov.vn/vn/tID32936_Gia-gao-Viet-Nam-tang-len-muc-cao-nhat-trong-hon-15-nam.html

AGROINFO

KẾT NỐI NGHIÊN CỨU VỚI THỰC TIỄN
CHO MỘT NỀN NÔNG NGHIỆP TĂNG TRƯỞNG TOÀN DIỆN

TRANG CHỦGIỚI THIỆU TIN TỨC

GIÁ GẠO VIỆT NAM TĂNG LÊN MỨC CAO NHẤT TRONG HƠN 15 NĂM

25 | 12 | 2023

Trong tuần này giá gạo đồ từ các nhà xuất khẩu hàng đầu đã tăng do nguồn cung hạn chế, đẩy giá gạo Việt Nam lên mức cao nhất trong hơn 15 năm và giá gạo của Ấn Độ lên mức cao nhất trong hai tháng.

Nguồn: doanh nghiệptiepthi.vn

Giá gạo đồ từ các nhà xuất khẩu hàng đầu đã tăng trong tuần này do nguồn cung hạn chế, đẩy giá gạo Việt Nam lên mức cao nhất trong hơn 15 năm và giá gạo của Ấn Độ lên mức cao nhất trong hai tháng.

Giá gạo 5% tấm của Việt Nam được chào bán ở mức 660-665 USD/tấn trong phiên 21/12, mức cao nhất kể từ giữa tháng 07/2008, so với mức 655-660 USD/tấn trong tuần trước. Mặc dù, giá gạo xuất khẩu của Việt Nam ngày 23/12/2023 có giảm nhưng vẫn ở mức cao.

Đại diện doanh nghiệp chuyên giao dịch về gạo xuất khẩu thành phố Hồ Chí Minh cho biết lý do giá gạo cao là nguồn cung trong nước và nước ngoài đều khan hiếm đã khiến giá gạo tăng cao.

Giá gạo đồ 5% tấm của Ấn Độ được chào bán ở mức 505-512 USD/tấn trong tuần này, mức cao nhất trong

Labeling Data

		<u>NHÂN ĐƯỢC GẮN</u>			
		0	1	2	3
<u>NHÂN TỔ TÁC ĐỘNG</u>	Biến động giá	Không có thông tin	Giảm	Ổn định	Tăng
	Thị trường		<ul style="list-style-type: none"> - Cung tăng - Cầu giảm - Cung vượt cầu 	Cung cầu cân bằng	<ul style="list-style-type: none"> - Cầu tăng - Cung giảm - Cầu vượt cung
	Chính sách		Đối nội	Đối ngoại	Khác
	Nội tại		Sản lượng	Chất lượng	Chi phí sản xuất
	Ngoại vi		Dịch bệnh	Thiên tai	Khủng hoảng

Tong, A. H. (2012). Factors influencing price of agricultural products and stability countermeasures. Asian agricultural research, 4(1812-2016-143301), 17-43.

Example

#	title	brief	date	content	sources	category	price	gov	market	intrinsic	extrinsic
0	VIỆT NAM ĐANG XUẤT NHIỀU GẠO NHẤT THẾ GIỚI	VN đang là nước xuất khẩu gạo nhiều và giá cao nhất thế giới. Chỉ tính riêng trong tháng 8, VN đã xuất khẩu gần 1 triệu tấn, mang về nguồn thu hơn nửa tỉ USD.	05-09-2023	Nguồn: thanhnien Từ ngày 20.7, Ấn Vượt Thái Lan cả Tính đến đầu thán VN đang tận dụng Công Hân	https:// agro. gov.vn /vn/tl D3250 2 -Vie	1	3	0	0	1	0
1	XUẤT KHẨU GẠO 8 THÁNG ĐẠT MỨC KỶ LỤC	8 tháng, xuất khẩu gạo đạt gần 6 triệu tấn, mức cao nhất từ trước đến nay, tăng 20% so cùng kỳ và hoàn thành 89% kế hoạch năm	12-09-2023	Nguồn: Vnexpress.net Đây là dữ liệu thống kê mới nhất từ Hải quan. Với mức xuất	https:// agro. gov.vn /vn/tl D3253 7 Xua	1	0	0	0	1	0
2	PHILIPPINES ĐÀM PHÁN MUA GẠO VIỆT NAM, ẤN ĐỘ	Đề hạ nhiệt trong nước, Philippines vừa nhận báo giá gạo tốt hơn từ Việt Nam và đang thuyết phục Ấn Độ bán hàng, theo Bloomberg	15-08-2023	Nguồn: Vnexpress.net Thứ trưởng Bộ Nông nghiệp Philippines Domingo	https:// agro. gov.vn /vn/tl D3243 8 Phil	1	3	3	3	0	0
3	GIỮ THƯƠNG HIỆU GẠO VIỆT GIỮA BIẾN ĐỘNG LƯƠNG THỰC TOÀN CẦU	Bộ trưởng Bộ Công Thương Nguyễn Hồng Diên nhiều lần nhấn mạnh đến sự hợp tác giữa các bộ, ngành, doanh nghiệp, địa phương sao cho việc tổ chức sản xuất lúa gạo thông suốt qua các chuỗi giá trị bền vững.	22-08-2023	Nguồn: Baohinhphu.vn Sau khi đạt mức xuất khẩu kỷ lục 7,1 triệu tấn gạo năm 2022, hoạt	https:// agro. gov.vn /vn/tl D3247 6 Giu	1	0	1	0	0	0

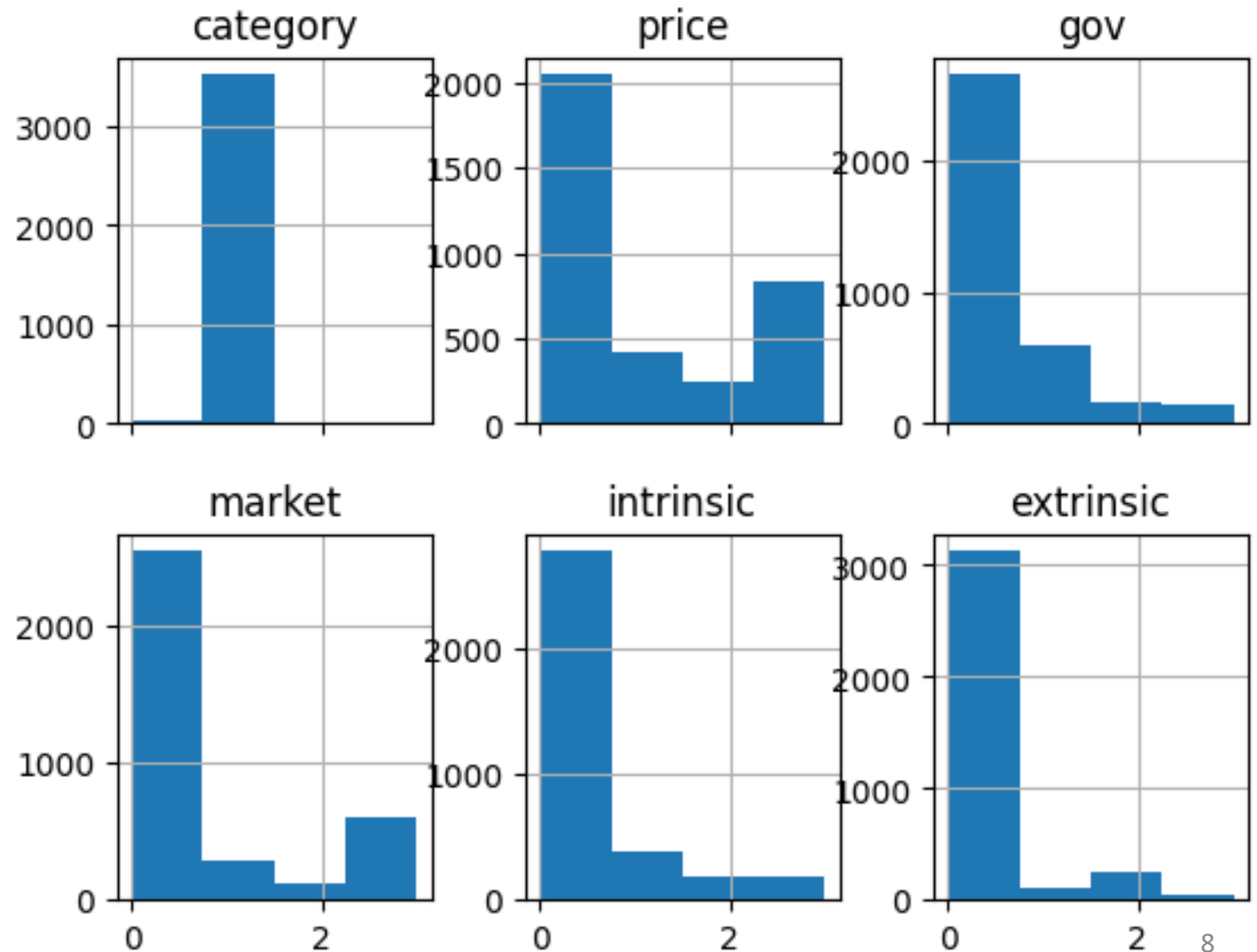
Data Distribution

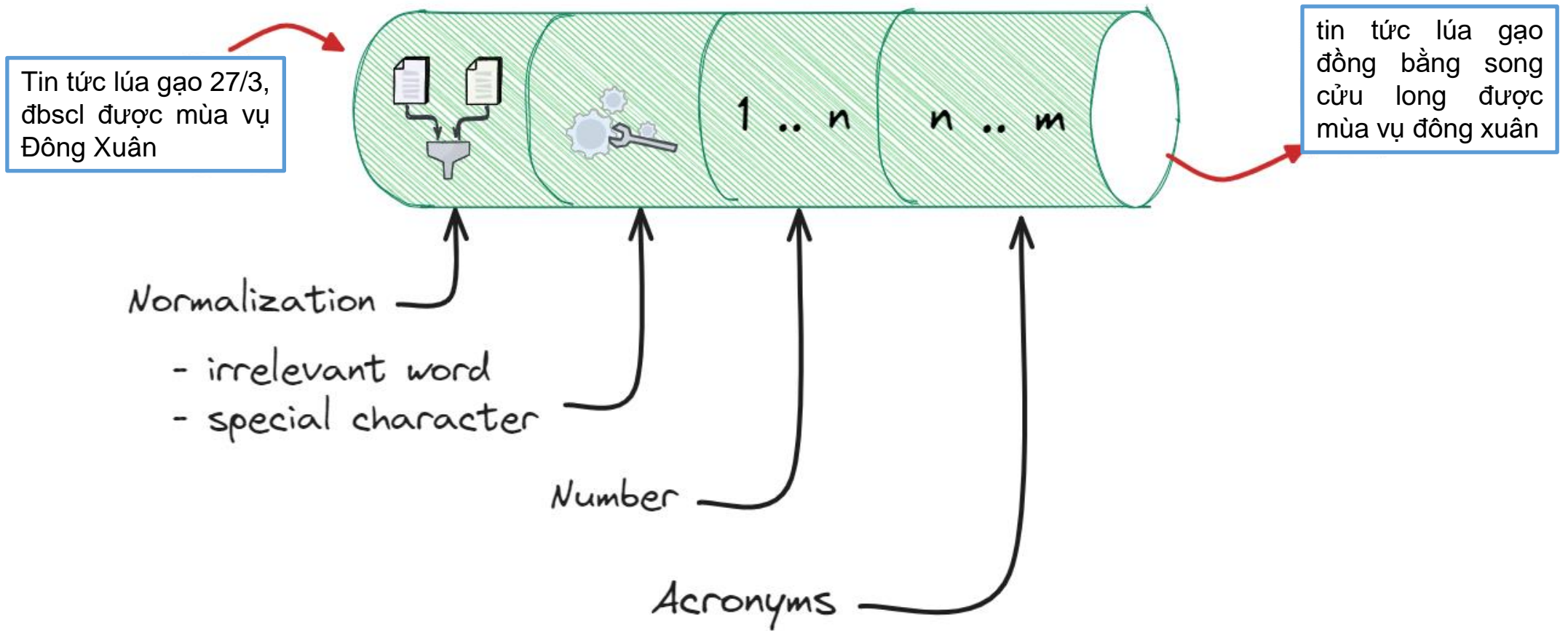
- Over **3554** samples

→ Small dataset

- **Imbalance** Data

→ Problem for training model





Data Processing Pipeline

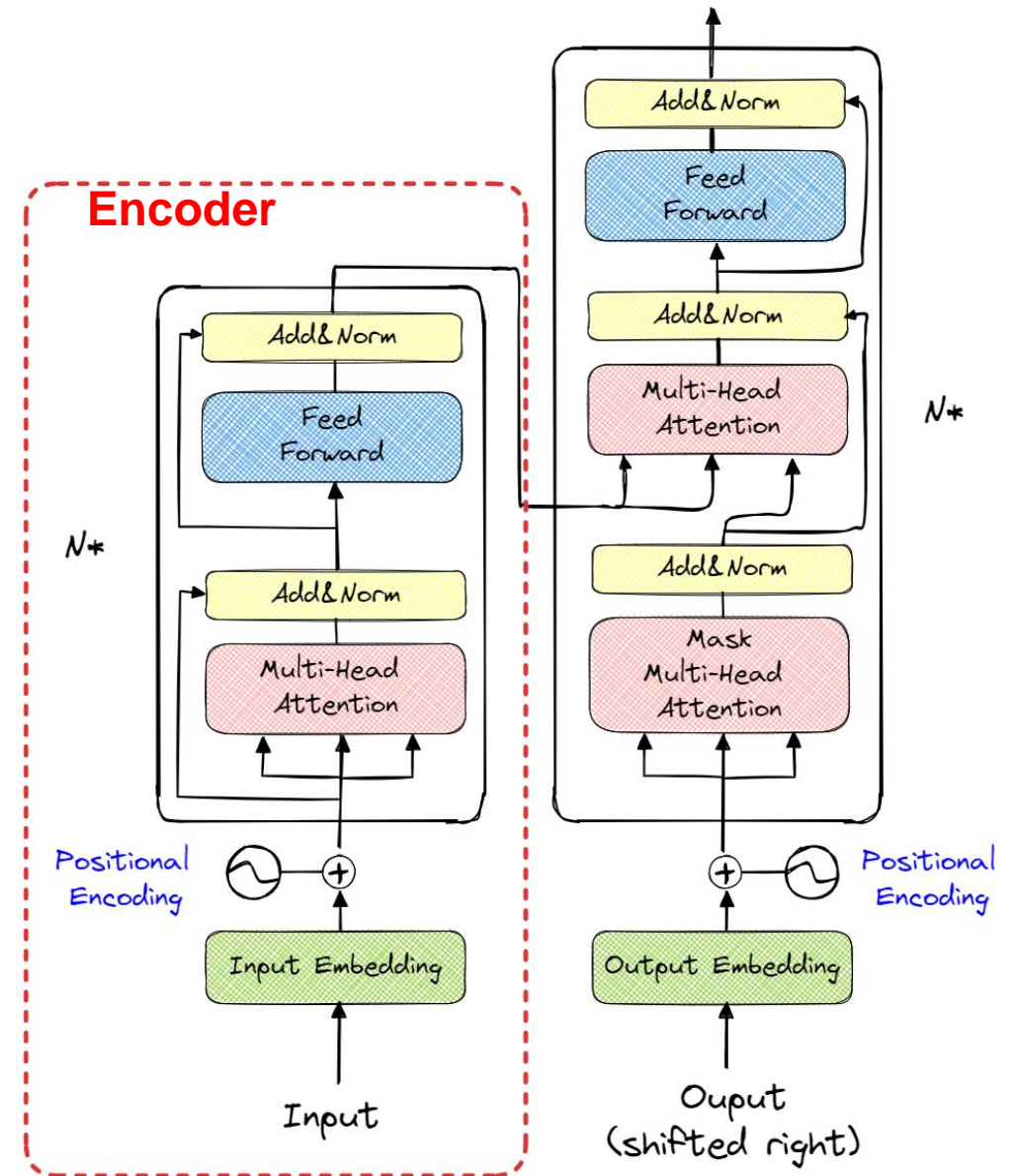
BERT

- **BERT** (Bidirectional Encoder Representations from Transformers) [1] is for pre-training Transformer's [2] encoder.
- **BERT** is highly efficient on many NLP tasks

Reference

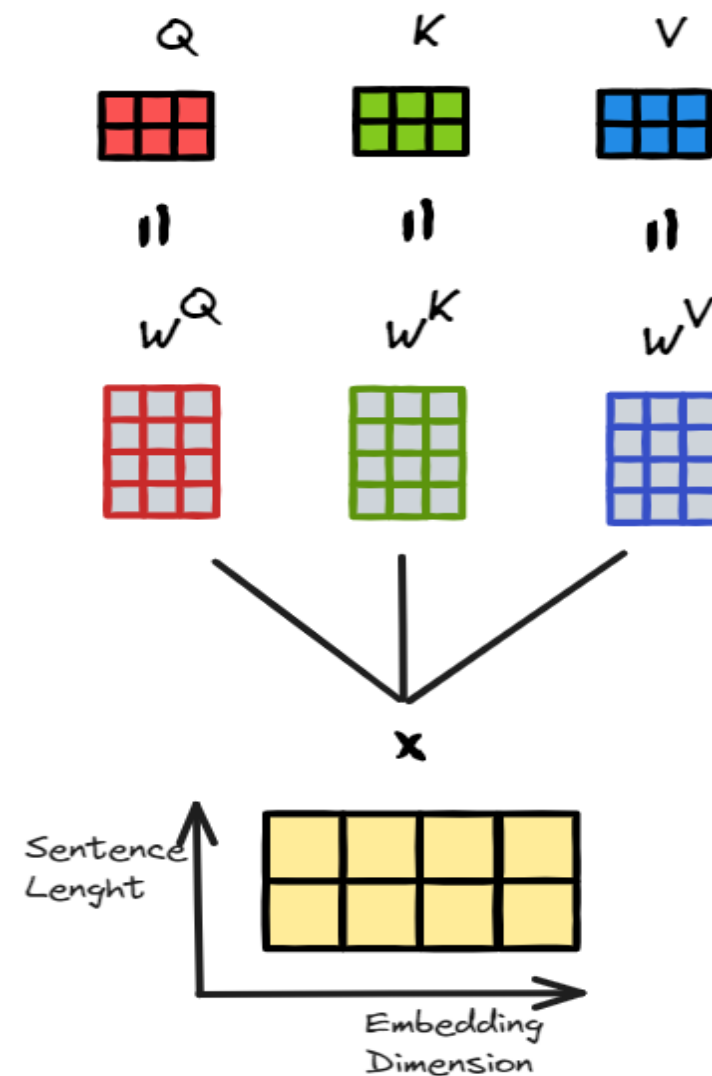
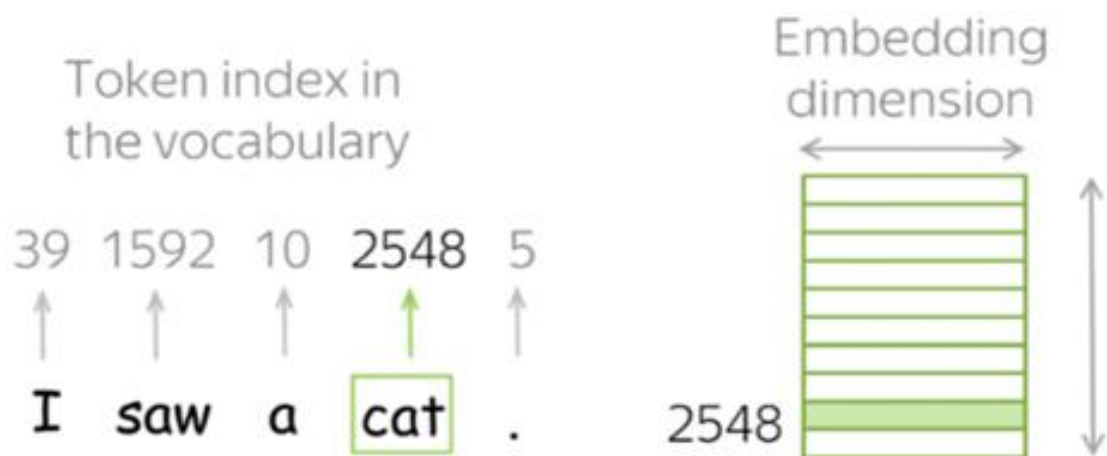
[1] Devlin, Chang, Lee, and Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In ACL, 2019.

[2] Vaswani and others. Attention is all you need. In NIPS, 2017



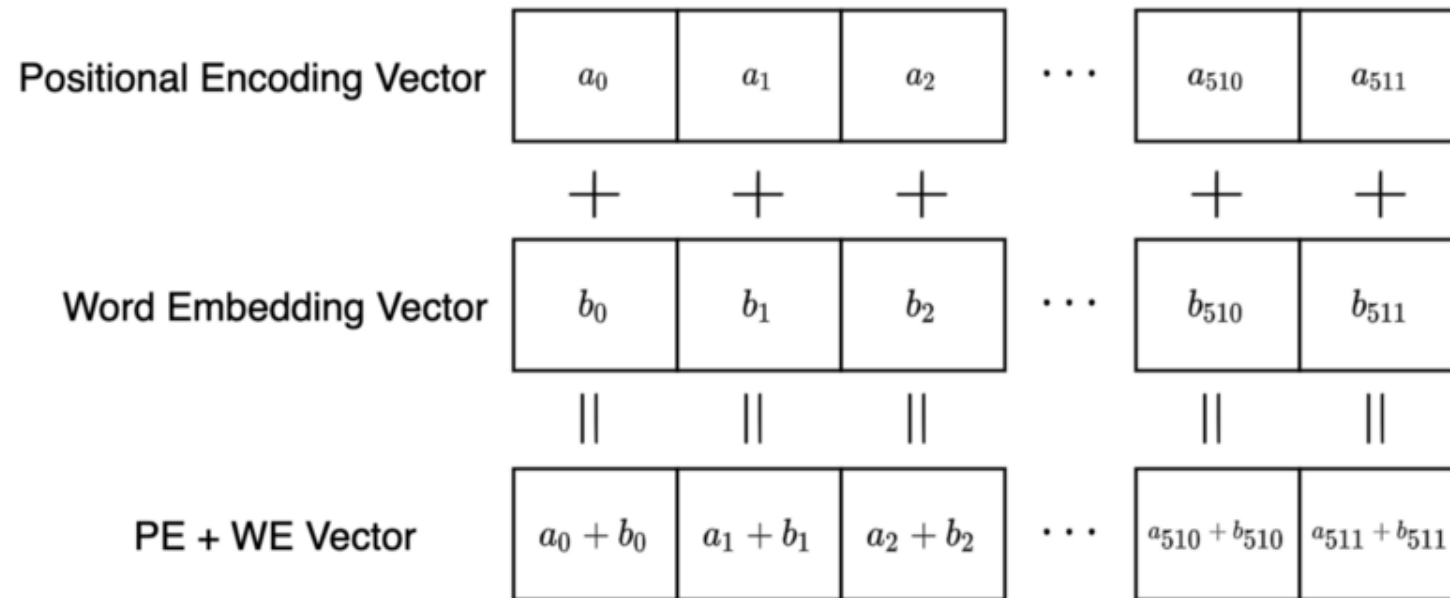
Input Embedding

- Machines cannot understand text, we need to convert them to vectors.
- In Transformer, the method used to embed text is the lookup table



Positional Encoding

For each word we need to add information about the position of that word through the Positional Encoding vector

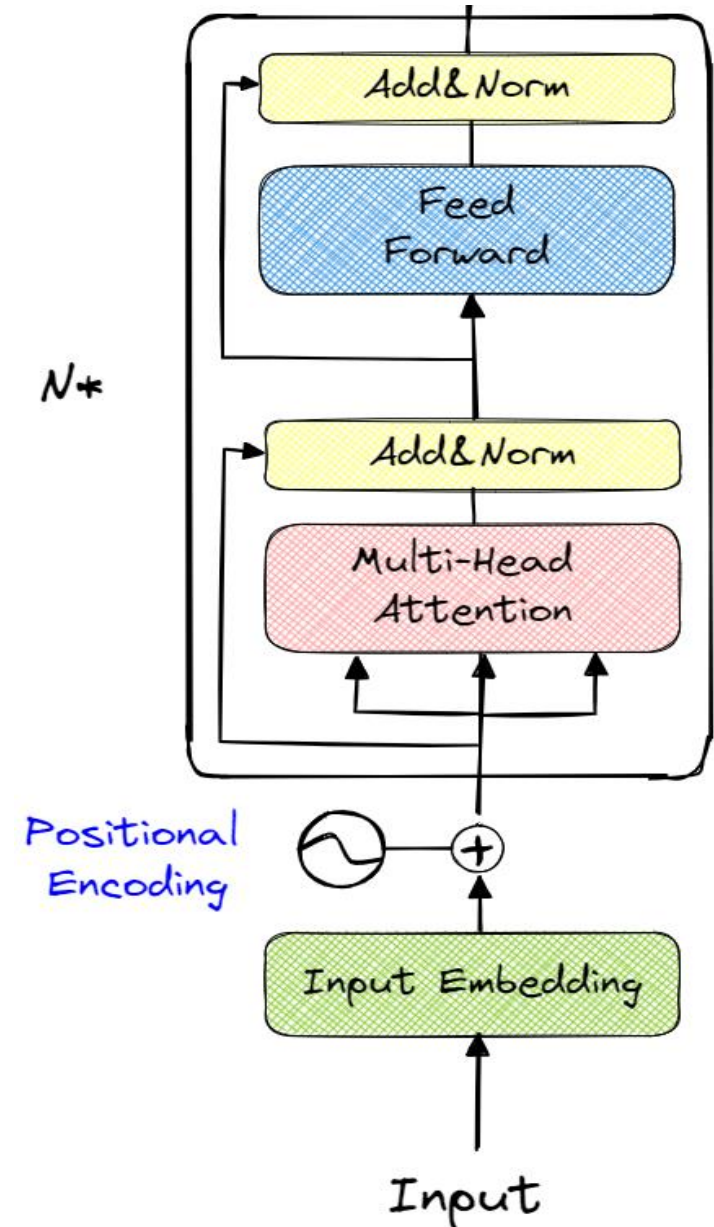


Encoder Block

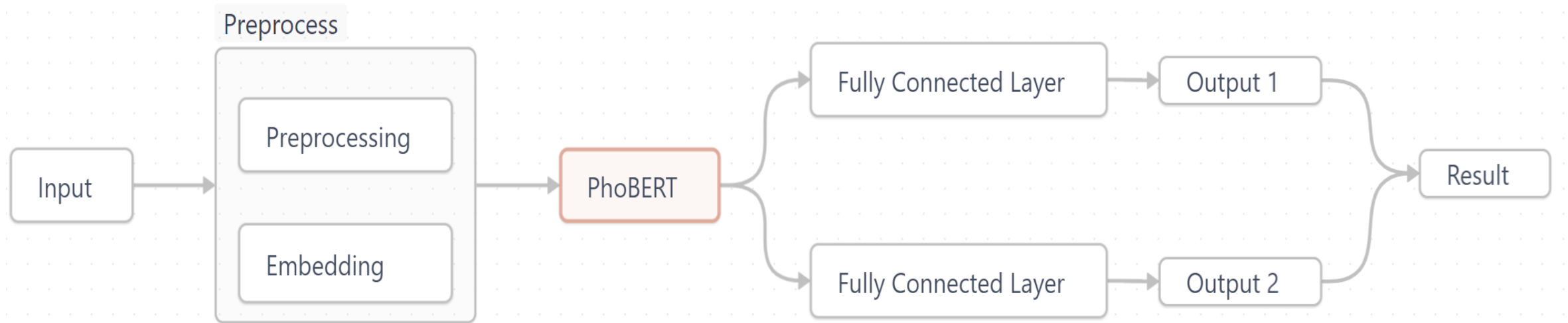
- **Multi-head attention** split Q, K, V multiple heads to calculate scale dot-product attention:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{Q \cdot K^T}{\sqrt{d}} \right) \cdot V$$

- **Feed forward** is a fully connected layer used to learn features from multi-head attention.



Pipeline



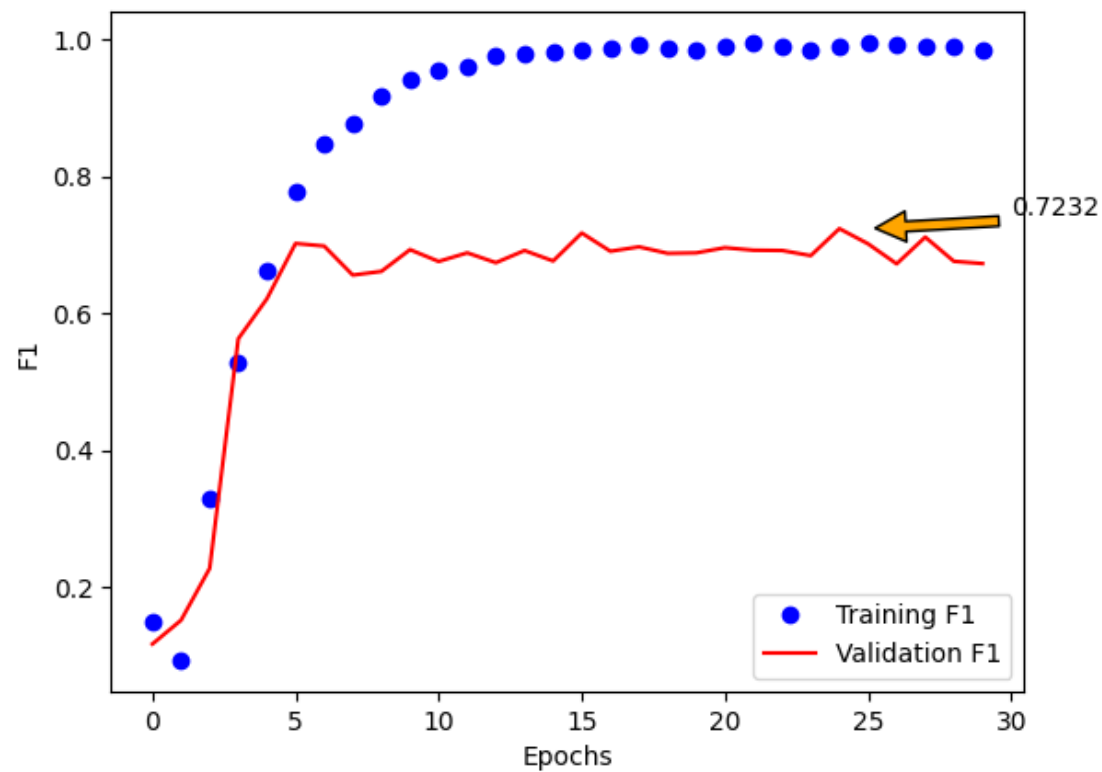
Loss Function & Metrics

- **Cross entropy** $= -\sum_{i=1}^{Class} p_i(\log q_i)$
- **Sigmoid Focal Loss** $= -\alpha(1 - p_t)^\gamma \log(p_t)$

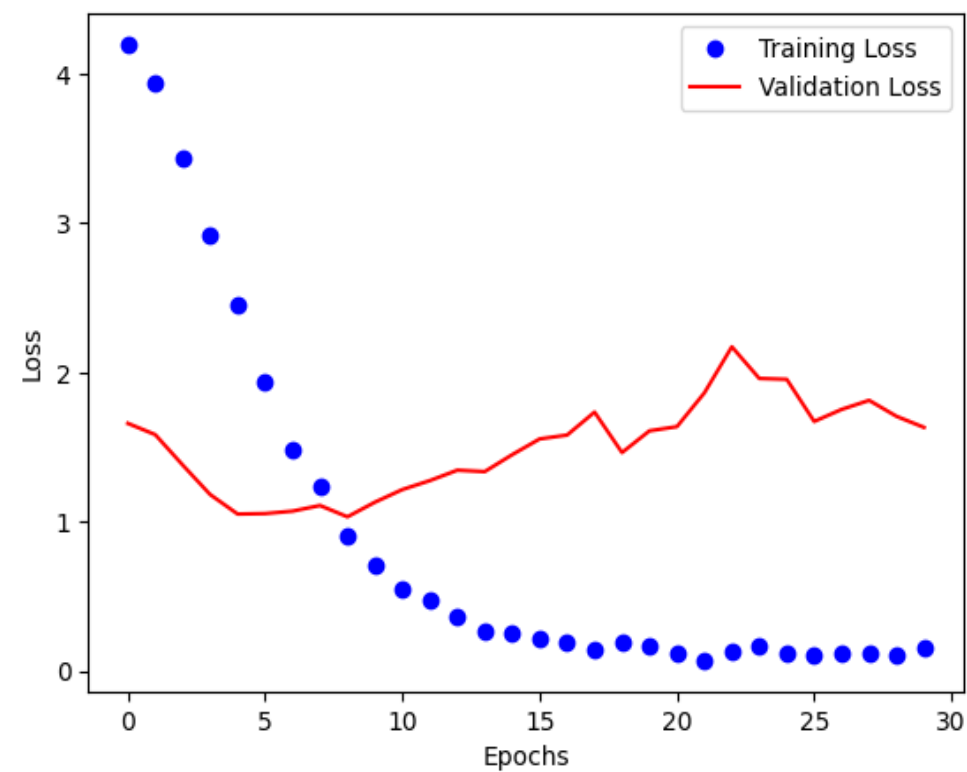
Loss function = Cross Entropy + Sigmoid Focal Loss

- $F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad \rightarrow \quad F_{macro} = \frac{\sum_i^N F_i}{N}$

Validation





Loss



Implement on Flask App

[TRANG CHỦ](#) [TIN TỨC ▾](#) [PHÂN TÍCH](#) [LIÊN HỆ](#)

Nhập Từ Khóa Để Tìm Kiếm  

PHÂN TÍCH TIN TỨC NÔNG SẢN

GIÁ XUẤT KHẨU TĂNG PHI MÃ, DOANH NGHIỆP XUẤT KHẨU G.

Giá gạo xuất khẩu của Việt Nam vẫn đang ở neo ở mức rất cao. Tuy nhiên, chỉ những doanh nghiệp bảo đảm được chuỗi cung ứng lúa gạo bền vững mới có thể tận dụng được tốt nhất cơ hội thị trường này.

I

[Phân Tích](#)

Giá Nông Sản:	Chưa có thông tin
Nông Sản:	Chưa có thông tin
Thị Trường:	Chưa có thông tin
Nội Tại:	Chưa có thông tin
Ngoại Vi:	Chưa có thông tin

Limitation

- Low F1-score
- Small labeled dataset



Development Direction

- Labeling more data
- Using ensemble techniques combined with time-series data



