```
import streamlit as st
import pandas as pd
import joblib
import re
from nltk.corpus import stopwords
from sklearn.metrics import accuracy_score

# Load saved preprocessing artifacts and model
vectorizer = joblib.load("vectorizer.pkl")
pca = joblib.load("pca.pkl")
svc = joblib.load("svm_model.pkl")
test_set = pd.read_csv("test_set.csv")  # Pre-saved test set for consistent accuracy

# Preprocessing function
def clean_text_column(df, column_name):
    stop_words = set(stopwords.words('english'))
    cleaned_texts = []

    for text in df[column_name]:
        if not isinstance(text, str):
            text = ""  # Handle non-string or NaN values
        text = text.replace('\n', '')  # Remove newlines
        words = re.findall(r'\b\w+\b', text.lower())  # Tokenize and lowercase
        filtered_words = [word for word in words if word not in stop_words]  # Remove
        cleaned_texts.append(' '.join(filtered_words))  # Join back into a string

    return cleaned_texts

# Header
st.title("News Aggregation and Filtering Tool")

# Display the image and team credit
st.image(
    "/Users/marwahfaraj/Desktop/ms_degree_application_and_doc/final_projects/502_fina
 # Replace with your image path
    use_column_width=True,
)

st.markdown(
    """
    <div style='text-align: center; font-size: 16px;'>
        <strong>Built by:</strong> Team 2 - Kirsten Drennen, Taylor Kirk, Marwah Fara
    </div>
```