DATA AND
ARTIFICIAL INTELLIGENCE

Capstone Session 11

simplilearn

DATA AND
ARTIFICIAL INTELLIGENCE

Deep Learning for Advanced Modeling

simplilearn

Deep Learning End Goal

The intuitive analyses of Aura must help customers make informed decisions to push relevant ads, services and products based on real-time user sentiments.

Aura's customers

Banking Organization

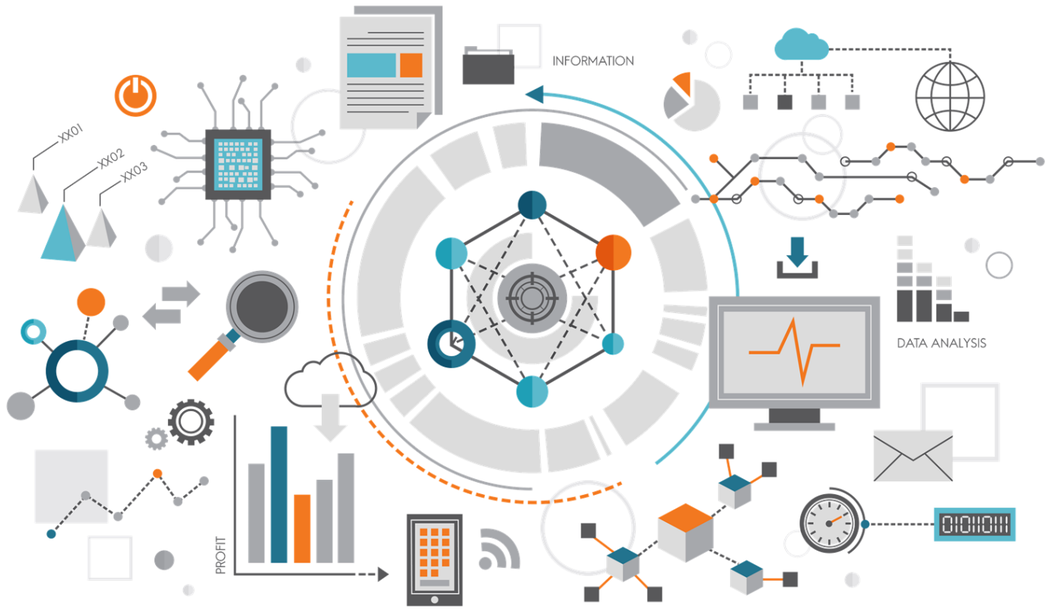
Cinema Owner

e-Commerce Company

Dental Clinic

Project Statement

Build necessary data aggregation, wrangling and visualization modules for Aura using the Healthcare dataset.



- Identify customers who churn the bank
- Detect humans wearing face masks
- Classify customer product reviews
- Denoise dirty documents

Week 11: Dataset Description

Variable	Description	Variable	Description
Id	Unique alphanumeric data	keys	Unique keys of the products
brand	Name of the brand	manufacturer	Name of the manufacturer
categories	Categories of the products Example: Food, Packaged food, Personal care	manufacturer Number	Number of the manufacturer
dateAdded	Details of date added	name	Name of the product Example: Lundberg Organic Cinnamon Toast Rice Cakes, Ambi Complexion Cleansing Bar
dateUpdated	Details of date updated	reviews.date	Date of the review
#ean	European Article Number	Reviews.date Added	Date on which the review was added

Week 11: Dataset Description

Variable	Description	Variable	Description
Reviews.dateSeen	Date on which the review was seen	reviews.text	Review of the product
Review.didPurchase	Binary data Example: True/False	reviews.title	Review title for the product Example: Good, Not worth it
reviews.doRecommend	Binary data Example: True/False	reviews.userCity	City of the user
Reviews.Id	Unique numeric data	reviews.userProvince	Province of the user
reviews.numHelpful	Number of reviews	reviews.username	Name of the user
reviews.rating	Number of rating	upc	Numeric data
reviews.sourceURLs	URL details	-	-

Week 11

Task: Build a CNN-LSTM hybrid model to classify the customer product reviews into good or bad.

Task A

- Load GrammarandProductReviews.csv dataset
- Create a feature named target by considering that a customer is pleased by the product if the rating is higher than 3. Any rating below 4 shows that the customer doesn't like the product. Use column reviews.rating to create feature target (Hint: df['target'] = df['reviews.rating'] < 4)
- Create your X with column reviews.text and Y with column target.
- Split your dataset into train and test in the ratio 80:20

Week 11

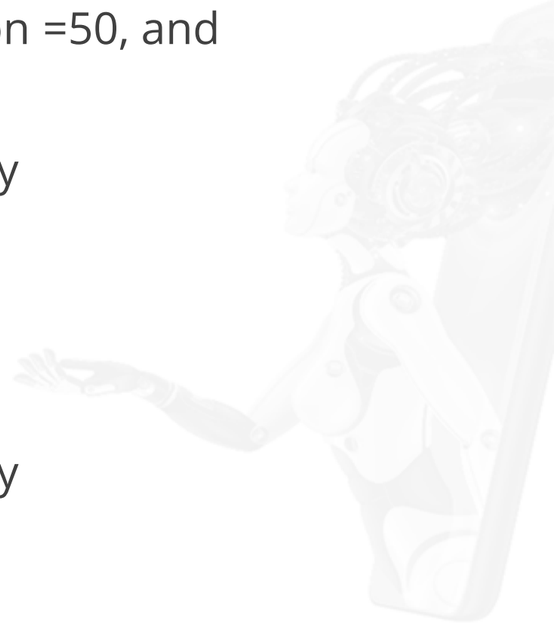
- Use a tokenizer from Keras to vectorize the text samples into a 2D integer tensor with 20000 words. Fit your tokenizer on train data. (MAX_NB_WORDS = 20000).

Note: You may use different tokenizers (from scikit-learn, NLTK, custom Python function etc.) This converts text into sequences of indices representing the 20000 most frequent words

- Convert train texts to sequences using the tokenizer texts_to_sequences method.
- Convert test texts to sequences using tokenizer texts_to_sequences method
- Pad train and test sequence (add 0s at the end until the sequence is of length 150). Consider MAX_SEQUENCE_LENGTH = 150 and this step gives your x_train and x_test
- One-hot encode your output classes (True/False).

Week 11

- Build a CNN - LSTM hybrid model with the following layers
  - Input layer with input shape = MAX_SEQUENCE_LENGTH and dtype int32
  - Embedding layer with input dimension = MAX_NB_WORDS, output dimension =50, and input length = MAX_SEQUENCE_LENGTH
  - Conv1D layer with 64 filters and kernel size 5 and activation relu, followed by MaxPooling1D with pool size = 5
  - Hint: MaxPool divides the length of the sequence by
  - Dropout(0.2)
  - Conv1D layer with 64 filters and kernel size 5 and activation relu, followed by MaxPooling1D with pool size = 5
  - Dropout(0.2)
  - LSTM layer with 64 units
  - Dense layer with 2 neurons and activation softmax



Week 11

- Compile the model with Adam optimizer and metric accuracy
- Train the model for 5 epochs and batch size 64
- Evaluate the model on test text print the test loss and accuracy

Task B

- As a future and take-home task, train the model with the full dataset available in this link - Grammar and Online Product Reviews | Kaggle.
- Evaluate the model on full test data and compare the performance improvement from a subset of the full dataset

DATA AND
ARTIFICIAL INTELLIGENCE

Thank You

simplilearn