

«*In The Name Of GOD*»



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

[HW-05-Report]

[NEURAL COMPUTING AND DEEP LEARNING]

Hasan Masroor | [403131030] | May 24, 2025

فهرست مطالب تمرین "05"

Question 1	2
1)	2
2)	3
3)	5
4)	8
5)	18
6)	20
7)	25

Problem 5: recurrent neural networks (RNN)

:Question 1 

1

در این تمرین، با استفاده از شبکه‌های عصبی بازگشتشی (RNN و LSTM) و مدل‌های ترکیبی (CNN+RNN)، ایمیل‌های فارسی را به دو دسته اسپم و عادی طبقه‌بندی می‌کنیم. مراحل کار شامل پیش‌پردازش داده‌ها، طراحی و مقایسه مدل‌های مختلف، تحلیل تأثیر پارامترهای کلیدی (مانند طول دنباله، اندازه تعبیه و ...) و ارزیابی عملکرد مدل‌ها با معیارهای Precision، Recall، Accuracy و F1-score است.

ابتدا کتابخانه‌های مورد نیاز را `import` می‌کنیم. ابتدا برای تحلیل بهتر ۵ نمونه اول دیتاست را نمایش می‌دهیم:

	text	label
0	...من پارسال اصلا آزاد شرکت نکن. ممنون آقا سامان	ham
1	...بالاخره آزمونارشد تموم شد من اسلام آقای کریمی	ham
2	... درود بر حاج وحیدی بنده بعنوان یک دکتری تاریخ	ham
3	... ضمن تقدیر از مسولین محترم اسلام و احترام	ham
4	... با سلام اینجانب یک دستگاه خودرو پراید ۱۳۱ با	ham

در ادامه، داده‌های متنی ایمیل‌های فارسی را برای آموزش مدل آماده کردیم. ابتدا متن‌ها را با حذف علائم نگارشی، اعداد، حروف لاتین و URLها پاک‌سازی نمودیم. سپس با استفاده از توکن‌ساز، متن‌ها را به کلمات مجزا تبدیل کرده و Stopwords رایج فارسی را حذف کردیم. در نهایت، برجسب‌های متنی (اسپم/عادی) را به مقادیر عددی (۰ و ۱) تبدیل و داده‌های پیش‌پردازش شده را ذخیره کردیم. این مراحل برای کاهش نویز و یکسان‌سازی ساختار متن قبل از ورود به مدل ضروری بودند. بعد از این کارها داده‌ها را به نسبت‌های ۷۰٪ برای آموزش، ۲۰٪ برای تست و ۱۰٪ برای اعتبارسنجی تقسیم کردیم. این تقسیم‌بندی با حفظ توزیع کلاس‌ها و به صورت تصادفی انجام شد تا مدل بتواند هم بر روی داده‌های آموزشی یاد بگیرد، هم بر روی داده‌های اعتبارسنجی تنظیم شود و در نهایت عملکرد نهایی آن بر روی داده‌های آزمون ارزیابی گردد.

در ادامه برای آماده‌سازی داده‌های متنی برای استفاده در مدل، ابتدا متن‌ها را با استفاده از توکن‌ساز به دنباله‌های عددی تبدیل کردیم و سپس این دنباله‌ها را به طول ثابت ۲۰۰ کلمه یکسان‌سازی کردیم. این کار را با حفظ قابلیت تکرارپذیری نتایج انجام دادیم تا مدل بهتر بتواند الگوهای موجود در داده‌ها را یاد بگیرد.

دیتاست شامل 1000 نمونه می‌باشد که بعد از تقسیم‌بندی به صورت زیر مجموعه‌ها را نمایش دادیم:

```
Train set: 700 samples
Validation set: 100 samples
Test set: 200 samples
```

.2

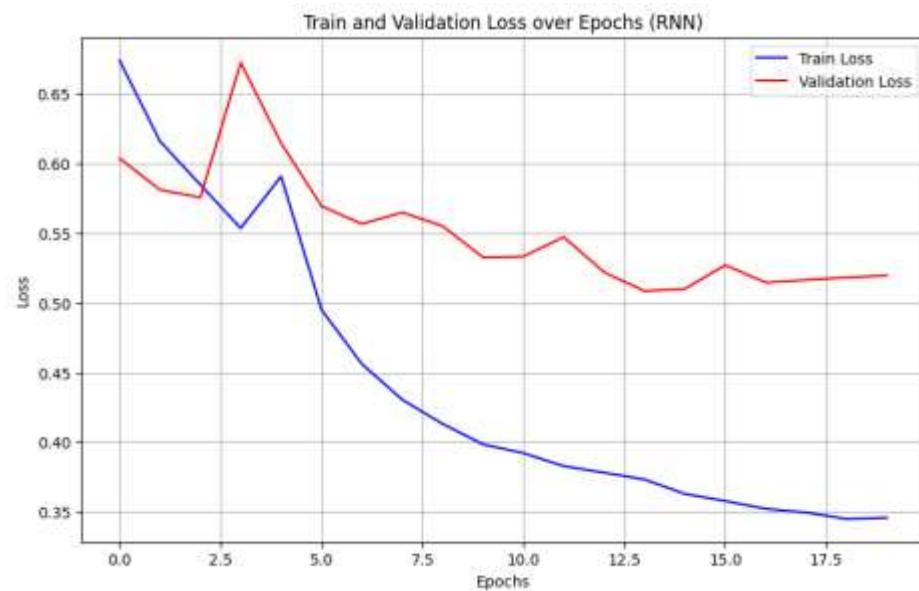
در این بخش باید یک مدل ساده متشکل از لایه تعبیه (Embedding)، لایه بازگشتی و لایه کاملاً متصل طراحی و آموزش دهیم، سپس عملکرد آن را با معیارهای ارزیابی ذکر شده نمایش دهیم و تحلیل کنیم. یک مدل ساده مبتنی بر شبکه عصبی بازگشتی (Simple RNN) طراحی و پیاده‌سازی کردیم. ابتدا پس از تعبیه توالی‌های متنی با استفاده از لایه Embedding، از یک لایه RNN با ۶۴ واحد برای استخراج ویژگی‌های زمانی استفاده کردیم. خروجی مدل با یک لایه Dense وتابع فعال‌سازی سیگموید برای انجام طبقه‌بندی دودویی نهایی شد. سپس مدل با استفاده از داده‌های آموزش، آموزش داده شده و عملکرد آن روی مجموعه تست ارزیابی کردیم. در نهایت خروجی آموزش این مدل و معیارهای ارزیابی محاسبه شده را در خروجی نمایش دادیم:

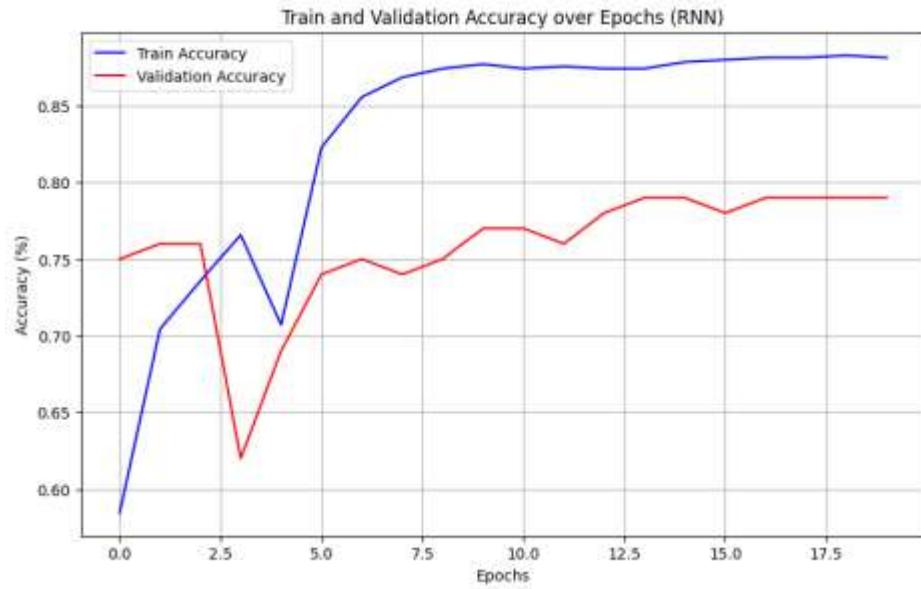
```
Epoch [1/20], Train Loss: 0.6737, Train Accuracy: 58.43%, Val Loss: 0.6035, Val Accuracy: 75.00%
Epoch [2/20], Train Loss: 0.6158, Train Accuracy: 70.43%, Val Loss: 0.5818, Val Accuracy: 76.00%
Epoch [3/20], Train Loss: 0.5846, Train Accuracy: 73.57%, Val Loss: 0.5754, Val Accuracy: 76.00%
Epoch [4/20], Train Loss: 0.5534, Train Accuracy: 76.57%, Val Loss: 0.6721, Val Accuracy: 62.00%
Epoch [5/20], Train Loss: 0.5906, Train Accuracy: 70.71%, Val Loss: 0.6144, Val Accuracy: 69.00%
Epoch [6/20], Train Loss: 0.4958, Train Accuracy: 82.29%, Val Loss: 0.5694, Val Accuracy: 74.00%
Epoch [7/20], Train Loss: 0.4561, Train Accuracy: 85.57%, Val Loss: 0.5564, Val Accuracy: 75.00%
Epoch [8/20], Train Loss: 0.4305, Train Accuracy: 86.86%, Val Loss: 0.5648, Val Accuracy: 74.00%
Epoch [9/20], Train Loss: 0.4131, Train Accuracy: 87.43%, Val Loss: 0.5550, Val Accuracy: 75.00%
Epoch [10/20], Train Loss: 0.3985, Train Accuracy: 87.71%, Val Loss: 0.5326, Val Accuracy: 77.00%
Epoch [11/20], Train Loss: 0.3922, Train Accuracy: 87.43%, Val Loss: 0.5332, Val Accuracy: 77.00%
Epoch [12/20], Train Loss: 0.3828, Train Accuracy: 87.57%, Val Loss: 0.5472, Val Accuracy: 76.00%
Epoch [13/20], Train Loss: 0.3781, Train Accuracy: 87.43%, Val Loss: 0.5219, Val Accuracy: 78.00%
Epoch [14/20], Train Loss: 0.3733, Train Accuracy: 87.43%, Val Loss: 0.5084, Val Accuracy: 79.00%
Epoch [15/20], Train Loss: 0.3629, Train Accuracy: 87.86%, Val Loss: 0.5099, Val Accuracy: 79.00%
Epoch [16/20], Train Loss: 0.3578, Train Accuracy: 88.00%, Val Loss: 0.5270, Val Accuracy: 78.00%
Epoch [17/20], Train Loss: 0.3522, Train Accuracy: 88.14%, Val Loss: 0.5145, Val Accuracy: 79.00%
Epoch [18/20], Train Loss: 0.3495, Train Accuracy: 88.14%, Val Loss: 0.5163, Val Accuracy: 79.00%
Epoch [19/20], Train Loss: 0.3449, Train Accuracy: 88.29%, Val Loss: 0.5181, Val Accuracy: 79.00%
Epoch [20/20], Train Loss: 0.3458, Train Accuracy: 88.14%, Val Loss: 0.5195, Val Accuracy: 79.00%
7/7 ━━━━━━ 0s 24ms/step

Final Test Accuracy: 73.50%
Precision: 91.2281
Recall: 52.0000
F1-Score: 66.2420
```

مدل در طول ۲۰ دوره آموزش، روند بهبودی در دقت آموزش (از ۵۸.۴۳ درصد به ۸۸.۱۴) و کاهش تدریجی خطای آموزش (از ۰.۳۴۵۸ به ۰.۶۷۳۷) را نشان می‌دهد، اما دقت اعتبارسنجی پس از رسیدن به ۷۹ درصد تقریباً ثابت مانده که حاکی از اشباع یادگیری و احتمالاً ظرفیت محدود مدل ساده RNN است. نکته قابل توجه در نتایج، فاصله معنادار میان دقت نهایی تست (۷۳.۵٪) و دقت اعتبارسنجی (۷۹٪) است که بیانگر عملکرد ضعیفتر مدل روی داده‌های خارج از مجموعه آموزش و در نتیجه محدودیت در تعمیم‌پذیری آن است. از سوی دیگر، مقدار Precision نسبتاً بالا (۹۱.۲٪) در کنار Recall پایین (۵۲٪) نشان می‌دهد که مدل تمایل دارد تنها نمونه‌های مثبت با اطمینان بالا را شناسایی کند و در نتیجه بسیاری از موارد واقعی را از دست می‌دهد.

سپس در ادامه نمودارهای خطای دقت را برای آموزش و اعتبارسنجی نمایش دادیم و نمودار پایین روند کاهش یکنواخت خطای آموزش را نشان می‌دهد، در حالی که خطای اعتبارسنجی پس از چند دوره نخست به یک سطح نسبتاً ثابت می‌رسد و در برخی نقاط نیز دچار نوسان می‌شود و این رفتار می‌تواند نشانه‌ای از بروز اورفیت در مدل باشد. در نمودار دقت نیز مشاهده می‌کنیم دقت آموزش به مرور افزایش پیدا کرده است اما در اعتبارسنجی می‌بینیم که در ابتدا دقت نوسان دارد و کاهش می‌یابد اما مجدد سیر صعودی به خود می‌گیرد. در ادامه می‌توانیم این نمودارها را مشاهده کنیم:





۳

در این مرحله، یک مدل LSTM طراحی کردیم تا ویژگی‌های زمانی موجود در توالی‌های متنه را بهتر درک کند. ابتدا با استفاده از لایه Embedding، ورودی‌های متنه را به dense vectors تبدیل کردیم و سپس آن‌ها را به لایه LSTM با ۶۴ واحد عصبی منتقل کردیم تا وابستگی‌های طولانی مدت بین کلمات استخراج شوند. در نهایت، مدل را با هدف طبقه‌بندی دودویی و با نرخ یادگیری ۰.۰۰۰۱ آموزش دادیم. در ادامه روند آموزش مدل را به صورت دقیق در هر epoch نشان دادیم و پس از اتمام آموزش، عملکرد نهایی مدل را از طریق پیش‌بینی روی داده‌های تست و محاسبه معیارهایی قبلی ارزیابی نمودیم.

سپس در ادامه خروجی این مدل را برای هر اپک نمایش دادیم. مدل LSTM در طی ۲۰ دوره آموزش، روندی پیوسته و مؤثر در کاهش خطای آموزش و افزایش دقت از خود نشان داده است؛ به طوری که دقت اعتبارسنجی از ۶۱٪ در ابتدا به ۹۶٪ در پایان دوره‌ها رسیده است. این بهبود چشم‌گیر نشان‌دهنده توانایی بالای مدل در یادگیری ویژگی‌های زمانی داده‌هاست. همچنین، دقت نهایی تست ۸۸.۵٪ در کنار Precision بسیار بالا (۹۸.۷۳٪) و Recall مناسب (۷۸٪) حاکی از تعادل نسبی بین اطمینان در پیش‌بینی و میزان پوشش نمونه‌های مثبت است. مقدار F1-Score معادل ۸۷.۱۵٪ نیز نشان می‌دهد که مدل توانسته عملکرد کلی قابل قبولی در دسته‌بندی دودویی ارائه دهد.

خروجی این مدل برای این 20 اپک و معیارهای محاسبه شده را می‌توانیم در شکل زیر مشاهده کنیم:

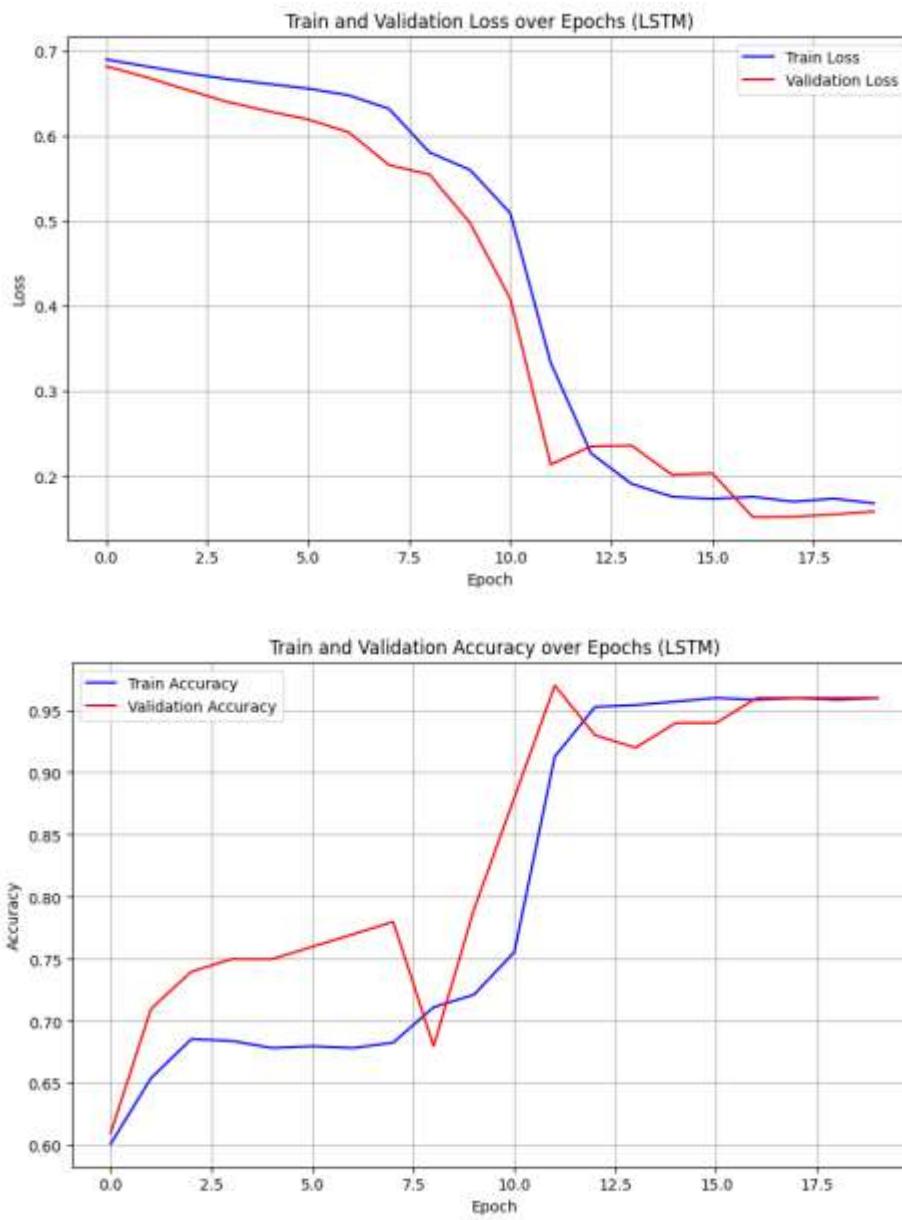
```

Epoch [1/20], Train Loss: 0.6900, Train Accuracy: 60.14%, Val Loss: 0.6820, Val Accuracy: 61.00%
Epoch [2/20], Train Loss: 0.6817, Train Accuracy: 65.43%, Val Loss: 0.6690, Val Accuracy: 71.00%
Epoch [3/20], Train Loss: 0.6739, Train Accuracy: 68.57%, Val Loss: 0.6543, Val Accuracy: 74.00%
Epoch [4/20], Train Loss: 0.6668, Train Accuracy: 68.43%, Val Loss: 0.6401, Val Accuracy: 75.00%
Epoch [5/20], Train Loss: 0.6613, Train Accuracy: 67.86%, Val Loss: 0.6293, Val Accuracy: 75.00%
Epoch [6/20], Train Loss: 0.6557, Train Accuracy: 68.00%, Val Loss: 0.6194, Val Accuracy: 76.00%
Epoch [7/20], Train Loss: 0.6478, Train Accuracy: 67.86%, Val Loss: 0.6044, Val Accuracy: 77.00%
Epoch [8/20], Train Loss: 0.6320, Train Accuracy: 68.29%, Val Loss: 0.5658, Val Accuracy: 78.00%
Epoch [9/20], Train Loss: 0.5807, Train Accuracy: 71.14%, Val Loss: 0.5548, Val Accuracy: 68.00%
Epoch [10/20], Train Loss: 0.5603, Train Accuracy: 72.14%, Val Loss: 0.4981, Val Accuracy: 79.00%
Epoch [11/20], Train Loss: 0.5094, Train Accuracy: 75.57%, Val Loss: 0.4086, Val Accuracy: 88.00%
Epoch [12/20], Train Loss: 0.3335, Train Accuracy: 91.29%, Val Loss: 0.2141, Val Accuracy: 97.00%
Epoch [13/20], Train Loss: 0.2271, Train Accuracy: 95.29%, Val Loss: 0.2353, Val Accuracy: 93.00%
Epoch [14/20], Train Loss: 0.1913, Train Accuracy: 95.43%, Val Loss: 0.2363, Val Accuracy: 92.00%
Epoch [15/20], Train Loss: 0.1762, Train Accuracy: 95.71%, Val Loss: 0.2020, Val Accuracy: 94.00%
Epoch [16/20], Train Loss: 0.1736, Train Accuracy: 96.00%, Val Loss: 0.2038, Val Accuracy: 94.00%
Epoch [17/20], Train Loss: 0.1761, Train Accuracy: 95.86%, Val Loss: 0.1523, Val Accuracy: 96.00%
Epoch [18/20], Train Loss: 0.1704, Train Accuracy: 96.00%, Val Loss: 0.1527, Val Accuracy: 96.00%
Epoch [19/20], Train Loss: 0.1739, Train Accuracy: 95.86%, Val Loss: 0.1553, Val Accuracy: 96.00%
Epoch [20/20], Train Loss: 0.1684, Train Accuracy: 96.00%, Val Loss: 0.1588, Val Accuracy: 96.00%
7/7 ━━━━━━ 0s 38ms/step

Final Test Metrics for LSTM:
Accuracy: 88.50%
Precision: 98.73%
Recall: 78.00%
F1-Score: 87.15%

```

سپس در ادامه نمودارهای خط و دقت را برای آموزش و اعتبارسنجی نمایش دادیم و نمودار پایین روند کاهش یکنواخت خطای آموزش را نشان می‌دهد، خطای اعتبارسنجی هم به همین صورت کاهش یکنواختی از خود نشان داده است. در نمودار دقت نیز مشاهده می‌کنیم دقت آموزش به مرور افزایش پیدا کرده است و از اپک 10 شیب صعودی زیادی می‌گیرد اما بعد از آن دقت تغییر آنچنانی نمی‌کند و ثابت می‌ماند. در اعتبارسنجی نیز می‌بینیم که در ابتدا دقت در حال افزایش است اما در اپک 7 به ناگهان دقت کاهش پیدا می‌کند اما از اپک 8 به بعد مجدد سیر صعودی با شیب نسبتاً زیادی می‌گیرد و در نهایت چهار نوسان های کاهش و افزایش می‌شود و مثل آموزش به شیب ثابتی می‌رسد. در ادامه می‌توانیم این نمودارها را مشاهده کنیم:



Final Test Metrics for LSTM:
Accuracy: 88.50%
Precision: 98.73%
Recall: 78.00%
F1-Score: 87.15%

Final Test Accuracy: 73.50%
Precision: 91.2281
Recall: 52.0000
F1-Score: 66.2420

شکل سمت راست برای RNN سوال 2 و سمت چپ نیز برای LSTM سوال 3 می باشد. می بینیم که دقت در این مدل افزایش قابل توجهی داشته است و نزدیک به 15 درصد زیاد شده است؛ همینطور برای سه معیار دیگر هم

می‌بینیم که افزایش خیلی خوبی نسبت به مدل سوال 2 داشته است. به طور کلی در این مدل توانستیم که معیارهای Precision، Accuracy، Recall و F1-score را بهبود بدهیم و مدل LSTM عملکرد خیلی بهتری از خود نشان داد و همینطور اگر به نمودارهای دقت و خطای برای آموزش و اعتبارسنجی نشان دادیم دقت کنیم، می‌بینیم که در مدل سوال 3 تا حد خوبی از نوسان‌های زیاد افزایش کاهش کاسته شده است و مدل روند خوبی را در پیش‌بینی ایمیل‌های اسپم و عادی در پیش‌گرفته است و با دقت کلی ۸۸.۵٪ به نتایج بهتری نسبت به مدل سوال دوم توانسته است دست یابد.

4

در این بخش باید با طراحی آزمایش‌های مختلف، تأثیر چهار پارامتر کلیدی مدل LSTM شامل طول دنباله ورودی، اندازه بردار تعبیه، تعداد واحدهای LSTM و استفاده یا عدم استفاده از Dropout را بر عملکرد مدل بررسی کنیم. برای هر پارامتر، حداقل سه مقدار مختلف را تست کرده و نتایج را به صورت نمودار و جدول ارائه می‌دهیم تا مشخص شود کدام پارامترها بیشترین تأثیر را معیارهای قبلی دارند.

در این مرحله، برای تحلیل حساسیت مدل نسبت به طول دنباله ورودی، طول‌های مختلفی از توالی (50، 100، 300 و 400) را مورد ارزیابی قرار دادیم، البته در سوال سوم مقدار 200 را داشتیم و پس داریم 5 مقدار مختلف را بررسی و مقایسه می‌کنیم. برای هر مقدار، داده‌ها را متناسب با طول موردنظر padding و truncating کردیم، سپس مدل را آموزش دادیم و در نهایت با محاسبه معیارهای عملکرد روی داده‌تست، تأثیر این پارامتر را بررسی کردیم و خروجی آن به شرح زیر است:

- **Input Sequence Length:**

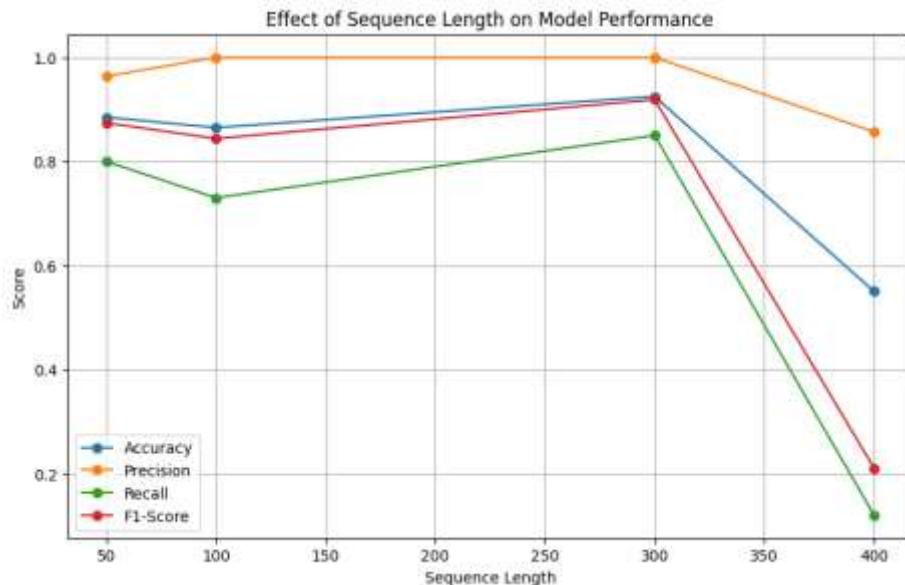
```
Evaluating for sequence length = 50
7/7 ━━━━━━ 0s 41ms/step
Accuracy: 0.8850 | Precision: 0.9639 | Recall: 0.8000 | F1-Score: 0.8743

Evaluating for sequence length = 100
7/7 ━━━━━━ 0s 40ms/step
Accuracy: 0.8650 | Precision: 1.0000 | Recall: 0.7300 | F1-Score: 0.8439

Evaluating for sequence length = 300
7/7 ━━━━━━ 1s 71ms/step
Accuracy: 0.9250 | Precision: 1.0000 | Recall: 0.8500 | F1-Score: 0.9189

Evaluating for sequence length = 400
7/7 ━━━━━━ 1s 81ms/step
Accuracy: 0.5500 | Precision: 0.8571 | Recall: 0.1200 | F1-Score: 0.2105
```

برای بررسی و تجزیه و تحلیل بهتر در ادامه خروجی های بالا را به صورت نمودار و جدول نمایش دادیم. تحلیل نتایج نشان می دهد که طول دنباله ورودی تأثیر قابل توجهی بر عملکرد مدل دارد. در ابتدا با افزایش طول از 50 به 100 دقیق کمی کاهش پیدا می کند اما با افزایش طول تا 300، شاهد بهبود چشمگیر در دقت، Recall و F1-score هستیم که بیانگر توانایی بهتر مدل در درک وابستگی های بلندمدت است. با این حال، افزایش بیش از حد طول (تا 400) منجر به افت شدید دقت و F1-score شده است، که احتمالاً ناشی از افزایش پیچیدگی و دشواری در یادگیری توالی های بلند است. در مجموع، طول 300 به نظر می رسد تعادلی مناسب میان پوشش اطلاعات و کارایی مدل ایجاد کرده است.



جدول خروجی این بخش نیز به صورت زیر است:

50	100	200	300	400	طول دنباله ورودی
88.5	86.5	88.5	92.5	55	Accuracy (%)
96.39	100	98.73	100	85.71	Precision (%)
80	73	78	85	12	Recall (%)
87.43	84.39	87.15	91.89	21.05	F1-score (%)

در این بخش، به بررسی تأثیر اندازه بردار تعییه (Embedding Dimension) بر عملکرد مدل پرداختیم. برای این منظور، مدل LSTM را با چهار مقدار مختلف برای ابعاد تعییه (32، 64، 128 و 200) آموخته دادیم، در حالی که سایر پارامترها ثابت نگه داشته شدند (مقدار 100 را نیز در سوال سوم در نظر گرفته بودیم و در نتیجه داریم 5 مقدار مختلف را بررسی می‌کنیم). با تغییر اندازه بردار، میزان ظرفیت مدل برای نمایش ویژگی‌های معنایی کلمات تغییر می‌کند. ابعاد بالاتر امکان دارد نمایش غنی‌تر و دقیق‌تری از واژگان را فراهم می‌کنند، اما ممکن است باعث افزایش پیچیدگی مدل و احتمال overfitting شوند. همچنین، عملکرد مدل در هر حالت با استفاده از معیارهای قبلی ارزیابی شدند تا تأثیر این پارامتر را بهتر بررسی کنیم. در نهایت خروجی حاصل از آموخته مدل روی مقادیر مختلف این پارامتر را در پایین نمایش دادیم:

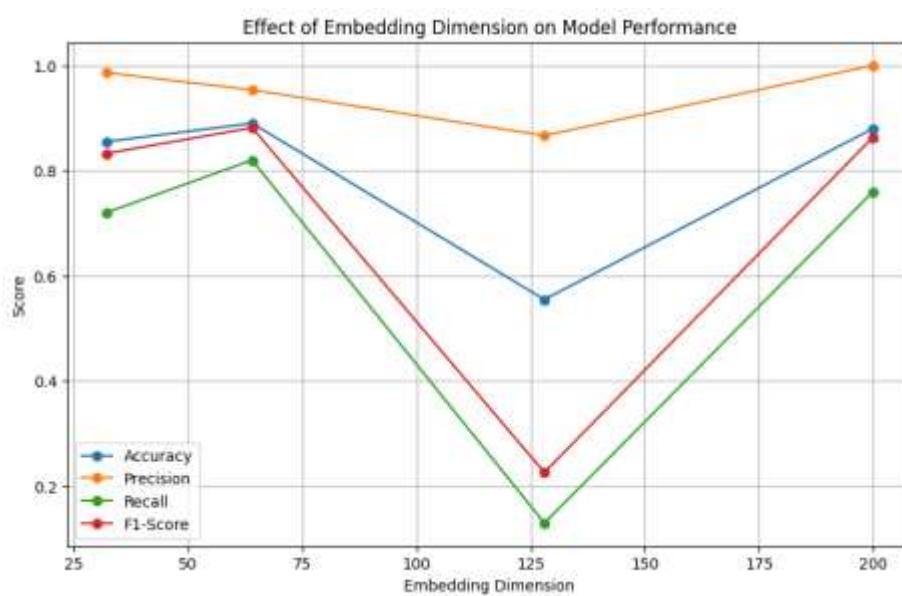
- **Embedding Vector Dimension:**

```
Evaluating for embedding_dim = 32
7/7 ━━━━━━━━ 1s 77ms/step
Accuracy: 0.8550 | Precision: 0.9863 | Recall: 0.7200 | F1-Score: 0.8324

Evaluating for embedding_dim = 64
7/7 ━━━━━━━━ 0s 50ms/step
Accuracy: 0.8900 | Precision: 0.9535 | Recall: 0.8200 | F1-Score: 0.8817

Evaluating for embedding_dim = 128
7/7 ━━━━━━━━ 1s 79ms/step
Accuracy: 0.5550 | Precision: 0.8667 | Recall: 0.1300 | F1-Score: 0.2261

Evaluating for embedding_dim = 200
7/7 ━━━━━━━━ 1s 78ms/step
Accuracy: 0.8800 | Precision: 1.0000 | Recall: 0.7600 | F1-Score: 0.8636
```



تحلیل نتایج نشان می‌دهد که اندازه بردار تعییه نقش مهمی در کیفیت یادگیری و عملکرد نهایی مدل دارد. در حالت کلی، افزایش ابعاد تعییه تا حد مشخصی (۶۴ و ۱۰۰) باعث بهبود پایداری مدل در تشخیص الگوهای معنایی شده و عملکرد بالای در تمامی معیارها بهویژه F1-score و Recall ایجاد کرده است. به طور خاص، اندازه ۶۴ با دقت ۸۹٪ و F1 با ۸۷٪ عملکردی متوازن از خود نشان داده است. با این حال، کاهش شدید در عملکرد مدل با بعد ۱۲۸ (بهویژه افت Recall به ۱۳٪ و F1-score به ۲۲.۶۱٪) نشان می‌دهد که افزایش ابعاد بدون تناسب با حجم داده و ساختار مدل می‌تواند موجب بیش برآش و ناپایداری در یادگیری شود. همچنین با بزرگتر شدن بُردار تا ۲۰۰، عملکرد مدل بهبود نسبی می‌یابد، اما همچنان نسبت به مقدار ۶۴، تفاوت محسوسی در Recall دارد. در مجموع، نتایج حاکی از آن است که انتخاب بهینه برای ابعاد تعییه نیازمند موازنیه‌ای دقیق میان قدرت نمایش و کنترل پیچیدگی مدل است و مقدار ۶۴ را در این آزمایشات می‌توانیم به عنوان یک مقدار بهینه در نظر بگیریم. جدول خروجی این بخش نیز به صورت زیر است:

اندازه بردار تعییه	32	64	100	128	200	
Accuracy (%)	85.5	89	88.5	55.5	88	
Precision (%)	98.63	95.35	98.73	86.67	100	
Recall (%)	72	82	78	13	76	
F1-score (%)	83.24	88.17	87.15	22.61	86.36	

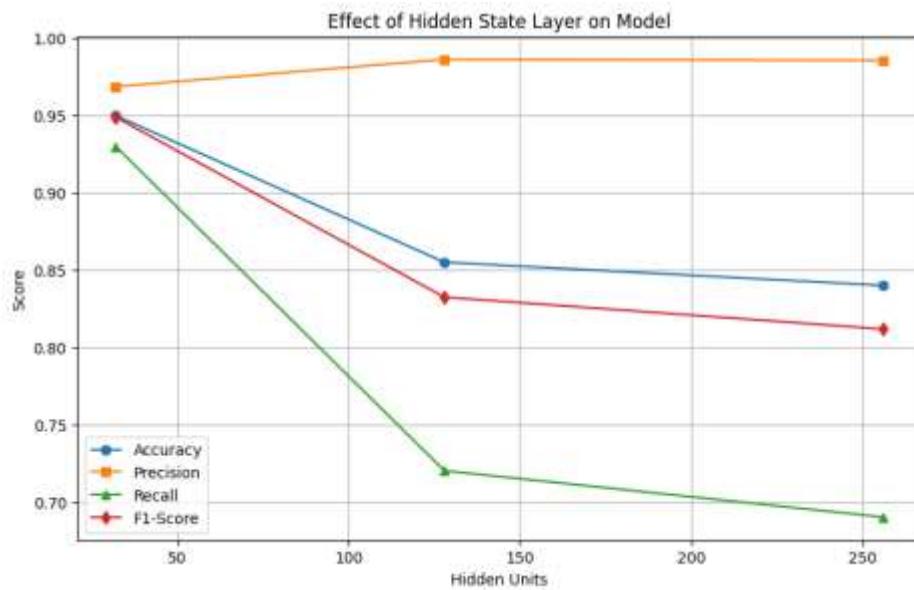
در بخش بعدی نیز به منظور بررسی تأثیر ابعاد بردار حالت (تعداد واحدهای LSTM) بر عملکرد مدل، مقدار این پارامتر را در سه سطح ۳۲، ۱۲۸ و ۲۵۶ آزمایش کردیم (مقدار ۶۴ را نیز در سوال سوم در نظر گرفته بودیم و در نتیجه داریم ۴ مقدار مختلف تعداد واحدهای LSTM، مدل ظرفیت بیشتری برای یادگیری وابستگی‌های پیچیده‌تر در توالی‌های متغیر پیدا می‌کند. با این حال این افزایش ممکن است منجر به بیش برآش یا افت عملکرد شود. ارزیابی نهایی نشان خواهد داد که کدام اندازه بردار حالت توافق نهایی میان دقت، یادگیری و تعمیم‌پذیری ایجاد می‌کند. همچنین، عملکرد مدل در هر حالت با استفاده از معیارهای قبلی ارزیابی شدند تا تأثیر این پارامتر را بهتر بررسی کنیم. در نهایت خروجی حاصل از آموزش مدل روی مقادیر مختلف این پارامتر را در پایین نمایش دادیم:

- **Hidden State layer of LSTM Network:**

```
Evaluating for LSTM hidden units = 32
7/7 ━━━━━━ 0s 28ms/step
Accuracy: 0.9500 | Precision: 0.9688 | Recall: 0.9300 | F1-Score: 0.9490

Evaluating for LSTM hidden units = 128
7/7 ━━━━━━ 1s 71ms/step
Accuracy: 0.8550 | Precision: 0.9863 | Recall: 0.7200 | F1-Score: 0.8324

Evaluating for LSTM hidden units = 256
7/7 ━━━━━━ 1s 135ms/step
Accuracy: 0.8400 | Precision: 0.9857 | Recall: 0.6900 | F1-Score: 0.8118
```



تحلیل نتایج نشان می‌دهد که ابعاد بردار حالت (تعداد واحدهای LSTM) تأثیر قابل توجهی بر توان مدل در یادگیری وابستگی‌های زمانی دارد. به طور کلی، با کاهش تدریجی اندازه بردار از ۲۵۶ به ۳۲، عملکرد مدل بهویژه در معیارهای Recall و F1-score بهبود چشمگیری پیدا کرده است. بردار حالت با اندازه ۳۲ با وجود سادگی، دقت ۹۴.۹٪ و F1-score معادل ۹۴.۹٪ را به دست آورده که نشان‌دهنده تعادل عالی بین دقت و توان تشخیص نمونه‌های مثبت است. در مقابل، افزایش بیش از حد تعداد واحدها (تا ۲۵۶) باعث افت نسبی در Recall (۸۱.۱٪) و F1-score (۸۶.۹٪) شده که ممکن است ناشی از بیش‌برازش یا موارد دیگر باشد. این نتایج حاکی از آن‌اند که مدل‌های ساده‌تر با بردارهای حالت کوچکتر، در این مسئله خاص عملکرد بهتری ارائه می‌دهند و ظرفیت بالا لزوماً به معنای کارایی بالاتر نیست.

جدول خروجی این بخش نیز به صورت زیر است و تجزیه و تحلیل های بالا را می توانیم در این جدول مشاهده و بررسی کنیم:

ابعاد بردار حالت	256	128	64	32
Accuracy (%)	84	85.5	88.5	95
Precision (%)	98.57	98.63	98.73	96.88
Recall (%)	69	72	78	93
F1-score (%)	81.18	83.24	87.15	94.9

در این بخش، به بررسی تأثیر استفاده از تکنیک Dropout بر عملکرد مدل می پردازیم. به عنوان یک روش Regularization نقش مهمی در جلوگیری از اورفیت ایفا می کند و می تواند به بهبود تعمیم پذیری مدل کمک کند. مدل بدون استفاده از Dropout را در سوال سوم استفاده کرده بودیم و در آن پارت کامل و دقیق موارد مختلف را بررسی کردیم و در ادامه توضیح مختصراً از نتایجی که در آن بخش بدون استفاده از Dropout بدست آوردیم می دهیم و پس از آن مدل با استفاده از Dropout را تعریف و آموزش می دهیم و سپس با حالت بدون Dropout مقایسه خواهیم کرد.

مدل LSTM در حالت بدون استفاده از Dropout، در طی 20 دوره آموزش روندی پیوسته و مؤثر در کاهش خطای آموزش و افزایش دقت از خود نشان داده است؛ به طوری که دقت اعتبارسنجی از 62٪ در ابتدا به 96٪ در پایان دوره‌ها رسیده است. این بهبود چشم‌گیر نشان‌دهنده توانایی بالای مدل در یادگیری ویژگی‌های زمانی داده‌هاست. همچنین، دقت نهایی تست 88٪ در کنار Precision بسیار بالا (98.73٪) و Recall مناسب (77٪) حاکی از تعادل نسبی بین اطمینان در پیش‌بینی و میزان پوشش نمونه‌های مثبت است. مقدار F1-Score معادل 86.15٪ نیز نشان می دهد که مدل توانسته عملکرد کلی قابل قبولی در دسته‌بندی دودویی ارائه دهد.

در ادامه نیز خروجی های این مدل در 20 اپک، معیارهای ارزیابی محاسبه شده و نمودارهای خطای آموزش و اعتبارسنجی را که در سوال 3 نمایش داده بودیم، مجدد برای یادآوری نمایش می دهیم که در نهایت مقایسه بهتری برای پارامتر Dropout داشته باشیم:

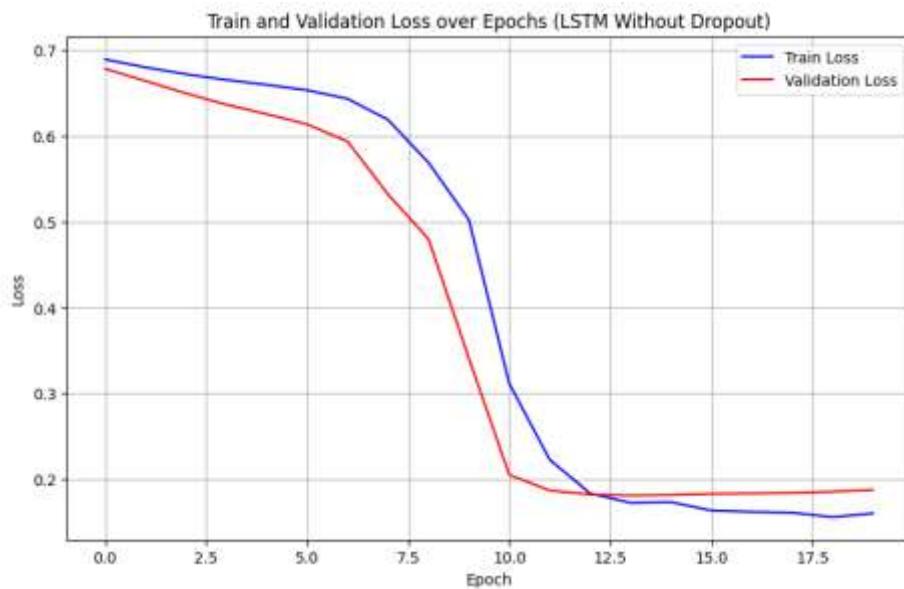
- Effect of Dropout Usage

- Without Dropout:

```
Epoch [1/20], Train Loss: 0.6890, Train Accuracy: 62.71%, Val Loss: 0.6780, Val Accuracy: 73.00%
Epoch [2/20], Train Loss: 0.6798, Train Accuracy: 66.29%, Val Loss: 0.6639, Val Accuracy: 74.00%
Epoch [3/20], Train Loss: 0.6719, Train Accuracy: 68.57%, Val Loss: 0.6495, Val Accuracy: 75.00%
Epoch [4/20], Train Loss: 0.6652, Train Accuracy: 68.14%, Val Loss: 0.6364, Val Accuracy: 75.00%
Epoch [5/20], Train Loss: 0.6596, Train Accuracy: 68.00%, Val Loss: 0.6253, Val Accuracy: 75.00%
Epoch [6/20], Train Loss: 0.6532, Train Accuracy: 68.00%, Val Loss: 0.6134, Val Accuracy: 77.00%
Epoch [7/20], Train Loss: 0.6432, Train Accuracy: 68.14%, Val Loss: 0.5933, Val Accuracy: 78.00%
Epoch [8/20], Train Loss: 0.6188, Train Accuracy: 68.57%, Val Loss: 0.5318, Val Accuracy: 78.00%
Epoch [9/20], Train Loss: 0.5689, Train Accuracy: 70.43%, Val Loss: 0.4798, Val Accuracy: 81.00%
Epoch [10/20], Train Loss: 0.5022, Train Accuracy: 77.29%, Val Loss: 0.3409, Val Accuracy: 93.00%
Epoch [11/20], Train Loss: 0.3115, Train Accuracy: 91.57%, Val Loss: 0.2052, Val Accuracy: 96.00%
Epoch [12/20], Train Loss: 0.2229, Train Accuracy: 95.00%, Val Loss: 0.1870, Val Accuracy: 95.00%
Epoch [13/20], Train Loss: 0.1836, Train Accuracy: 96.00%, Val Loss: 0.1823, Val Accuracy: 95.00%
Epoch [14/20], Train Loss: 0.1726, Train Accuracy: 96.00%, Val Loss: 0.1810, Val Accuracy: 95.00%
Epoch [15/20], Train Loss: 0.1734, Train Accuracy: 96.00%, Val Loss: 0.1817, Val Accuracy: 95.00%
Epoch [16/20], Train Loss: 0.1636, Train Accuracy: 96.29%, Val Loss: 0.1830, Val Accuracy: 95.00%
Epoch [17/20], Train Loss: 0.1621, Train Accuracy: 96.29%, Val Loss: 0.1835, Val Accuracy: 95.00%
Epoch [18/20], Train Loss: 0.1610, Train Accuracy: 96.29%, Val Loss: 0.1842, Val Accuracy: 95.00%
Epoch [19/20], Train Loss: 0.1560, Train Accuracy: 96.43%, Val Loss: 0.1856, Val Accuracy: 95.00%
Epoch [20/20], Train Loss: 0.1602, Train Accuracy: 96.29%, Val Loss: 0.1878, Val Accuracy: 95.00%
7/7 ━━━━━━ 1s Sims/step
```

Final Test Metrics for LSTM:

Accuracy: 88.00%
Precision: 98.72%
Recall: 77.00%
F1-Score: 86.52%





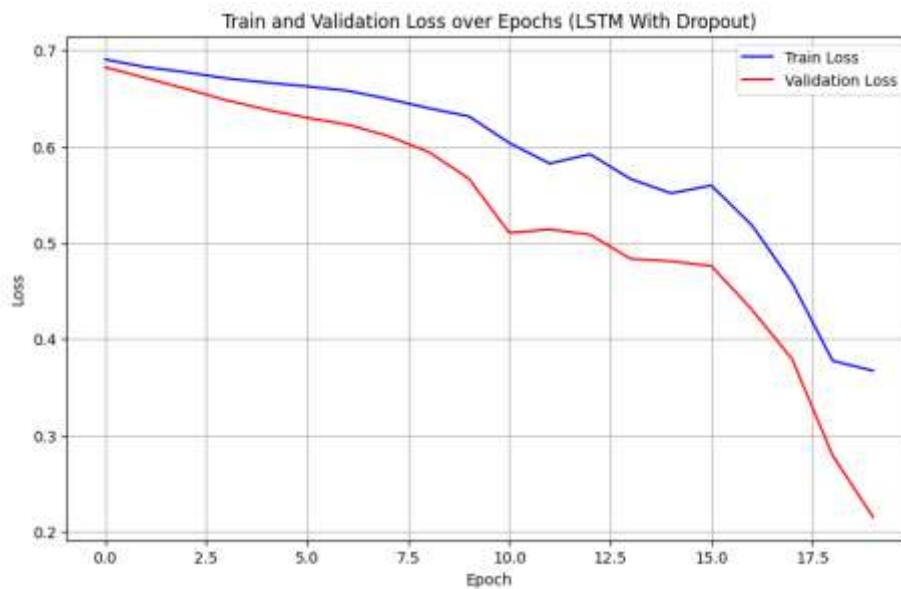
در ادامه نیز یک مدل LSTM با استفاده از Dropout و Recurrent Dropout پیاده‌سازی کردیم تا از بیش‌برازش در فرآیند آموزش جلوگیری شود. لایه LSTM با ۶۴ واحد حافظه و نرخ Dropout برابر ۰.۳ برای هر دو نوع تنظیم شده است. هدف از این پیاده‌سازی، مقایسه عملکرد آن با مدل بدون dropout و بررسی تأثیر منظم‌سازی بر تعمیم‌پذیری مدل در داده‌های آزمون است.

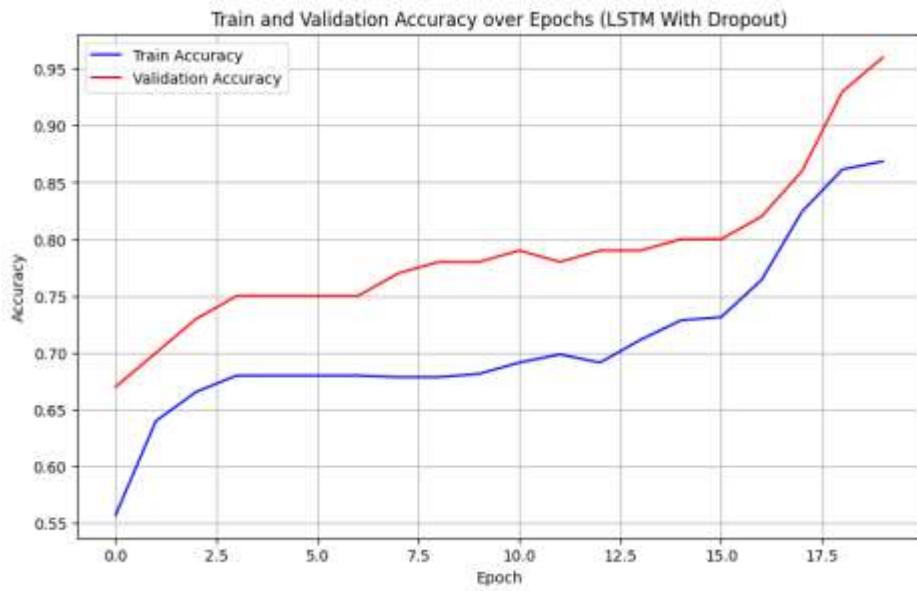
در ادامه خروجی برای هر اپک را نمایش دادیم و نتایج نشان می‌دهند که افزودن Dropout به لایه LSTM منجر به پایداری بیشتر یادگیری و بهبود چشمگیر در عملکرد تعمیم‌پذیری مدل شده است. برخلاف مدل ساده، این مدل با وجود افزایش تدریجی دقت آموزش، افزایش مداوم و همزمان در دقت اعتبارسنجی نیز نشان می‌دهد (از ۶۷٪ به ۹۶٪)، که بیانگر اثربخشی منظم‌سازی در جلوگیری از بیش‌برازش است. همچنین، دقت نهایی تست معادل ۹۱.۵٪، همراه با Precision بالا (۹۸.۸۲٪) و Recall قابل توجه (۸۴٪)، نشان‌دهنده توانایی بالای مدل در تشخیص نمونه‌های مثبت و منفی با تعادل مناسب است. در مجموع، استفاده از Dropout باعث شده تا مدل نه تنها در آموزش بهتر عمل کند، بلکه روی داده‌های نادیده نیز عملکرد قابل اطمینان داشته باشد و نسبت به مدل بدون دراپ‌اوت، بهبود محسوسی در F1-score و دقت نهایی نشان دهد.

- With Dropout:

```
Epoch [1/20], Train Loss: 0.6905, Train Accuracy: 55.71%, Val Loss: 0.6824, Val Accuracy: 67.00%
Epoch [2/20], Train Loss: 0.6825, Train Accuracy: 64.00%, Val Loss: 0.6712, Val Accuracy: 70.00%
Epoch [3/20], Train Loss: 0.6772, Train Accuracy: 66.57%, Val Loss: 0.6600, Val Accuracy: 73.00%
Epoch [4/20], Train Loss: 0.6708, Train Accuracy: 68.00%, Val Loss: 0.6481, Val Accuracy: 75.00%
Epoch [5/20], Train Loss: 0.6664, Train Accuracy: 68.00%, Val Loss: 0.6383, Val Accuracy: 75.00%
Epoch [6/20], Train Loss: 0.6625, Train Accuracy: 68.00%, Val Loss: 0.6299, Val Accuracy: 75.00%
Epoch [7/20], Train Loss: 0.6580, Train Accuracy: 68.00%, Val Loss: 0.6227, Val Accuracy: 75.00%
Epoch [8/20], Train Loss: 0.6495, Train Accuracy: 67.86%, Val Loss: 0.6110, Val Accuracy: 77.00%
Epoch [9/20], Train Loss: 0.6398, Train Accuracy: 67.86%, Val Loss: 0.5945, Val Accuracy: 78.00%
Epoch [10/20], Train Loss: 0.6314, Train Accuracy: 68.14%, Val Loss: 0.5667, Val Accuracy: 78.00%
Epoch [11/20], Train Loss: 0.6037, Train Accuracy: 69.14%, Val Loss: 0.5103, Val Accuracy: 79.00%
Epoch [12/20], Train Loss: 0.5823, Train Accuracy: 69.86%, Val Loss: 0.5138, Val Accuracy: 78.00%
Epoch [13/20], Train Loss: 0.5919, Train Accuracy: 69.14%, Val Loss: 0.5085, Val Accuracy: 79.00%
Epoch [14/20], Train Loss: 0.5664, Train Accuracy: 71.14%, Val Loss: 0.4834, Val Accuracy: 79.00%
Epoch [15/20], Train Loss: 0.5514, Train Accuracy: 72.86%, Val Loss: 0.4808, Val Accuracy: 80.00%
Epoch [16/20], Train Loss: 0.5595, Train Accuracy: 73.14%, Val Loss: 0.4759, Val Accuracy: 80.00%
Epoch [17/20], Train Loss: 0.5184, Train Accuracy: 76.43%, Val Loss: 0.4387, Val Accuracy: 82.00%
Epoch [18/20], Train Loss: 0.4585, Train Accuracy: 82.43%, Val Loss: 0.3790, Val Accuracy: 86.00%
Epoch [19/20], Train Loss: 0.3774, Train Accuracy: 86.14%, Val Loss: 0.2791, Val Accuracy: 93.00%
Epoch [20/20], Train Loss: 0.3671, Train Accuracy: 86.86%, Val Loss: 0.2150, Val Accuracy: 96.00%
7/7 ━━━━━━ 1s 66ms/step
```

```
Final Test Metrics for LSTM with Dropout:
Accuracy: 91.56%
Precision: 98.82%
Recall: 84.00%
F1-Score: 90.81%
```





نمودارهای مربوط به دقت و خطا در دوره‌های آموزش مدل LSTM با دراپ‌اوت، روندی بسیار امیدوارکننده را نشان می‌دهند. در نمودار دقت، مشاهده می‌شود که دقت اعتبارسنجی به‌طور پیوسته و با شیب مثبت تا بیش از ۹۵٪ افزایش یافته که حاکی از اثربخشی دراپ‌اوت در کاهش بیش‌بازارش و بهبود تعیین‌پذیری مدل است. در حالی که دقت آموزش رشد ملایم‌تری دارد، این اختلاف منطقی بوده و نشان از کنترل یادگیری بیش از حد بر داده‌های آموزش دارد. نمودار خطا نیز کاهش یکنواخت در هر دو مجموعه آموزش و اعتبارسنجی را نشان می‌دهد، به‌ویژه افت محسوس خطای اعتبارسنجی در دوره‌های پایانی که بیانگر ثبات یادگیری مدل و همگرایی مؤثر آن است. این روندها به خوبی مؤید تاثیر مثبت دراپ‌اوت در بهبود پایداری و عملکرد کلی مدل هستند.

Final Test Metrics for LSTM with Dropout:
 Accuracy: 91.50%
 Precision: 98.82%
 Recall: 84.00%
 F1-Score: 90.81%

Final Test Metrics for LSTM:
 Accuracy: 88.00%
 Precision: 98.72%
 Recall: 77.00%
 F1-Score: 86.52%

شکل سمت راست برای مدل بدون استفاده از Dropout است و سمت چپ نیز برای مدل با استفاده از Dropout می‌باشد. می‌بینیم که دقت در این مدل افزایش خوبی داشته است و نزدیک به ۳.۵ درصد زیاد شده است؛ همینطور برای سه معیار دیگر هم می‌بینیم که افزایش نسبت به مدل بدون Dropout داشته است. به طور کلی در این مدل توانستیم که معیارهای Accuracy، Precision، Recall و F1-score را بهبود بدهیم و مدل LSTM عملکرد خیلی بهتری از خود نشان داد و همینطور اگر به نمودارهای دقت و خطا که برای آموزش و اعتبارسنجی نشان دادیم دقت کنیم، می‌بینیم که در این مدل تا حد خیلی خوبی از اورفیت جلوگیری کرده است و مدل روند خوبی را در پیش‌بینی ایمیل‌های اسپم و عادی در پیش‌گرفته است و با دقت کلی ۹۱.۵٪ به نتایج بهتری نسبت به مدل بدون Dropout توانسته است دست یابد.

۵

در این بخش، با بهره‌گیری از ساختار دوطرفه (Bidirectional LSTM) برای لایه LSTM، سعی داریم عملکرد مدل را در درک بهتر وابستگی‌های زمانی بهبود دهیم. مدل‌های Bidirectional با پردازش داده‌ها در هر دو جهت، قادرند اطلاعات متنی را با دقت بالاتری تحلیل کرده و روابط معنایی عمیق‌تری میان واژگان استخراج کنند. این رویکرد به‌ویژه در مسائل طبقه‌بندی متون که توالی واژگان نقش کلیدی دارند، می‌تواند منجر به بهبود قابل توجه در دقت و تعمیم‌پذیری مدل شود.

خروجی این مدل برای این 20 اپک و معیارهای محاسبه شده را می‌توانیم در شکل زیر مشاهده کنیم:

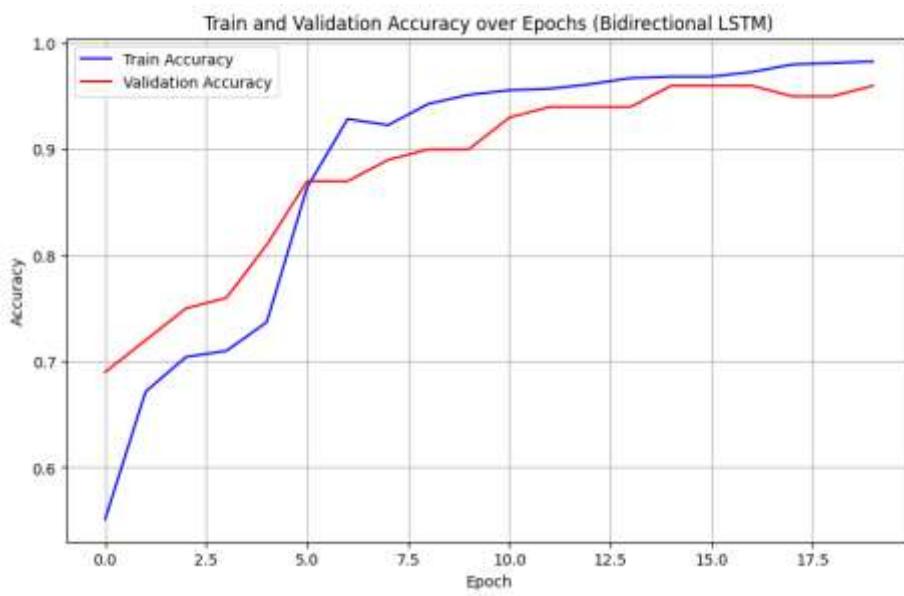
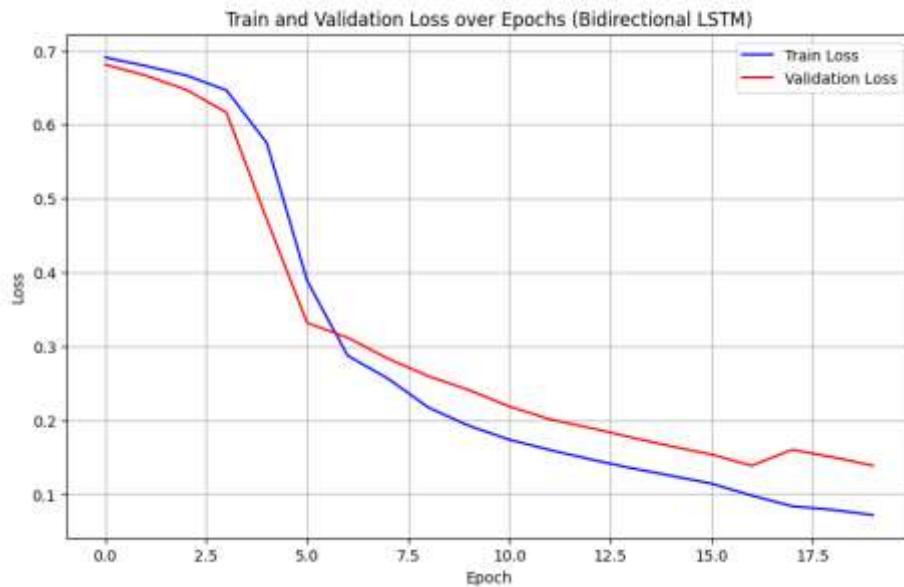
```
Epoch [1/20], Train Loss: 0.6909, Train Accuracy: 55.14%, Val Loss: 0.6812, Val Accuracy: 69.00%
Epoch [2/20], Train Loss: 0.6794, Train Accuracy: 67.14%, Val Loss: 0.6664, Val Accuracy: 72.00%
Epoch [3/20], Train Loss: 0.6667, Train Accuracy: 70.43%, Val Loss: 0.6474, Val Accuracy: 75.00%
Epoch [4/20], Train Loss: 0.6465, Train Accuracy: 71.00%, Val Loss: 0.6166, Val Accuracy: 76.00%
Epoch [5/20], Train Loss: 0.5751, Train Accuracy: 73.71%, Val Loss: 0.4716, Val Accuracy: 81.00%
Epoch [6/20], Train Loss: 0.3884, Train Accuracy: 86.43%, Val Loss: 0.3316, Val Accuracy: 87.00%
Epoch [7/20], Train Loss: 0.2873, Train Accuracy: 92.86%, Val Loss: 0.3117, Val Accuracy: 87.00%
Epoch [8/20], Train Loss: 0.2562, Train Accuracy: 92.29%, Val Loss: 0.2833, Val Accuracy: 89.00%
Epoch [9/20], Train Loss: 0.2171, Train Accuracy: 94.29%, Val Loss: 0.2597, Val Accuracy: 96.00%
Epoch [10/20], Train Loss: 0.1924, Train Accuracy: 95.14%, Val Loss: 0.2408, Val Accuracy: 98.00%
Epoch [11/20], Train Loss: 0.1736, Train Accuracy: 95.57%, Val Loss: 0.2186, Val Accuracy: 93.00%
Epoch [12/20], Train Loss: 0.1597, Train Accuracy: 95.71%, Val Loss: 0.2012, Val Accuracy: 94.00%
Epoch [13/20], Train Loss: 0.1471, Train Accuracy: 96.14%, Val Loss: 0.1896, Val Accuracy: 94.00%
Epoch [14/20], Train Loss: 0.1353, Train Accuracy: 96.71%, Val Loss: 0.1770, Val Accuracy: 94.00%
Epoch [15/20], Train Loss: 0.1248, Train Accuracy: 96.86%, Val Loss: 0.1647, Val Accuracy: 96.00%
Epoch [16/20], Train Loss: 0.1142, Train Accuracy: 96.86%, Val Loss: 0.1537, Val Accuracy: 96.00%
Epoch [17/20], Train Loss: 0.0979, Train Accuracy: 97.29%, Val Loss: 0.1384, Val Accuracy: 96.00%
Epoch [18/20], Train Loss: 0.0835, Train Accuracy: 98.00%, Val Loss: 0.1601, Val Accuracy: 95.00%
Epoch [19/20], Train Loss: 0.0786, Train Accuracy: 98.14%, Val Loss: 0.1499, Val Accuracy: 95.00%
Epoch [20/20], Train Loss: 0.0716, Train Accuracy: 98.29%, Val Loss: 0.1385, Val Accuracy: 96.00%
7/7 → 1s 71ms/step

Test Metrics for Bidirectional LSTM:
Accuracy: 95.50%
Precision: 98.92%
Recall: 92.00%
F1-Score: 95.34%
```

مدل Bidirectional LSTM در مقایسه با مدل LSTM یک‌طرفه، بهبود چشمگیری در عملکرد نشان داده است. دقت تست مدل دوطرفه به ۹۵.۵۰٪ رسیده، در حالی که مدل یک‌طرفه ساده به حدود ۸۸٪ محدود شده بود. همچنین کاهش یکنواخت و پایدار در مقدار Loss و افزایش پیوسته در دقت اعتبارسنجی، حاکی از یادگیری عمیق‌تر و استفاده بهتر از اطلاعات متنی در هر دو جهت است. مقدار F1-Score بازخوانی مناسی (Recall = ۹۲٪) برابر با ۹۵.۳۴٪ نشان می‌دهد که مدل نه تنها دقت بالا (Precision = ۹۸.۹۲٪) داشته و برخلاف نسخه یک‌طرفه، توانسته توازن مؤثری بین این دو معیار برقرار کند. این نتایج مؤید قدرت مدل‌های دوچهتی در استخراج ویژگی‌های زمینه‌ای غنی‌تر در داده‌های متنی هستند.

با رسم نمودارهای خط و دقت برای آموزش و اعتبارسنجی نیز می توانیم تحلیل های بالا را بهتر درک کنیم و می بینیم که خطای هر دو با شبی خوبی در حال کاهش هست و مثل مدل یک طرفه زیاد دچار اورفیت نشده است؛ همچنین برای دقت هم می بینیم که هر دو نمودار با شبی معادلی در حال افزایش هستند و به طور کلی بهبود و افزایش دقت و ... بهتری نسبت به مدل یک طرفه از خود نشان داده است.

در ادامه نمودارهای خط و دقت را نیز می توانیم برای درک بهتر مشاهده کنیم:



Test Metrics for Bidirectional
Accuracy: 95.50%
Precision: 98.92%
Recall: 92.00%
F1-Score: 95.34%

Final Test Metrics for LSTM:
Accuracy: 88.50%
Precision: 98.73%
Recall: 78.00%
F1-Score: 87.15%

Final Test Metrics for LSTM with Dropout:
Accuracy: 91.50%
Precision: 98.82%
Recall: 84.00%
F1-Score: 90.81%

شکل سمت چپ برای دوطرفه، وسط برای همان مدل LSTM ساده سوال سوم و سمت راستی با اعمال Dropout است و می‌بینیم که مدل دوطرفه از هردوی آنها عملکرد بهتری در هر ۴ معیار ارزیابی بررسی شده دارد و حرف‌ها و تحلیل‌های بالای ما درسا است و این مدل عمل و بهبود بهتری را ارخود نشان داده است.

6

در این بخش، هدف طراحی یک مدل ترکیبی است که از مزایای هم‌زمان لایه‌های پیچشی یک‌بعدی برای استخراج ویژگی‌های محلی و لایه LSTM برای درک وابستگی‌های زمانی در داده‌های متغیر بهره می‌برد. ترکیب این دو نوع لایه به مدل امکان می‌دهد تا ابتدا الگوهای معنایی کوتاه‌مدت را شناسایی کرده و سپس روابط طولانی‌مدت را مدل‌سازی کند. در ادامه، تأثیر تعداد لایه‌های CNN+Pooling بر عملکرد نهایی مدل بررسی و با ساختارهای پیشین مقایسه خواهد شد.

در این پیاده‌سازی، مدلی ترکیبی از CNN و LSTM طراحی می‌کنیم که ابتدا با استفاده از یک لایه تعییه، ورودی‌های متغیر را به بردارهای عددی تبدیل می‌کنیم. سپس با بهره‌گیری از یک لایه پیچشی یک‌بعدی و لایه‌ی زیرنمونه‌برداری، ویژگی‌های محلی را استخراج و فشرده می‌کنیم. در ادامه، این ویژگی‌ها را به یک لایه LSTM وارد می‌کنیم تا وابستگی‌های زمانی را مدل‌سازی کنیم و در نهایت، خروجی مدل را از طریق یک لایه چگال دودوی پیش‌بینی می‌کنیم. سپس مدل را به مدت ۲۰ دوره آموزشی آموخته می‌دهیم و مقادیر دقت و خطای را در هر دوره ثبت و نمایش می‌دهیم. در پایان نیز عملکرد مدل را با استفاده از معیارهای ارزیابی گفته شده می‌سنجدیم.

خروجی این مدل برای این ۲۰ اپک و معیارهای محاسبه شده را می‌توانیم در شکل زیر مشاهده کنیم:

- 1-layer:

```

Epoch [1/20], Train Loss: 0.6883, Train Accuracy: 65.71%, Val Loss: 0.6782, Val Accuracy: 78.00%
Epoch [2/20], Train Loss: 0.6792, Train Accuracy: 68.00%, Val Loss: 0.6629, Val Accuracy: 78.00%
Epoch [3/20], Train Loss: 0.6663, Train Accuracy: 68.14%, Val Loss: 0.6365, Val Accuracy: 78.00%
Epoch [4/20], Train Loss: 0.6463, Train Accuracy: 67.86%, Val Loss: 0.5958, Val Accuracy: 78.00%
Epoch [5/20], Train Loss: 0.6144, Train Accuracy: 69.00%, Val Loss: 0.5284, Val Accuracy: 78.00%
Epoch [6/20], Train Loss: 0.5581, Train Accuracy: 71.43%, Val Loss: 0.4599, Val Accuracy: 79.00%
Epoch [7/20], Train Loss: 0.4836, Train Accuracy: 74.29%, Val Loss: 0.3648, Val Accuracy: 86.00%
Epoch [8/20], Train Loss: 0.2974, Train Accuracy: 89.57%, Val Loss: 0.1784, Val Accuracy: 95.00%
Epoch [9/20], Train Loss: 0.2009, Train Accuracy: 95.00%, Val Loss: 0.3360, Val Accuracy: 87.00%
Epoch [10/20], Train Loss: 0.1894, Train Accuracy: 94.86%, Val Loss: 0.2662, Val Accuracy: 91.00%
Epoch [11/20], Train Loss: 0.1521, Train Accuracy: 96.29%, Val Loss: 0.1369, Val Accuracy: 96.00%
Epoch [12/20], Train Loss: 0.1594, Train Accuracy: 96.29%, Val Loss: 0.1373, Val Accuracy: 96.00%
Epoch [13/20], Train Loss: 0.1657, Train Accuracy: 95.71%, Val Loss: 0.3834, Val Accuracy: 87.00%
Epoch [14/20], Train Loss: 0.2126, Train Accuracy: 94.00%, Val Loss: 0.3741, Val Accuracy: 87.00%
Epoch [15/20], Train Loss: 0.1769, Train Accuracy: 95.29%, Val Loss: 0.3316, Val Accuracy: 88.00%
Epoch [16/20], Train Loss: 0.1605, Train Accuracy: 95.86%, Val Loss: 0.3271, Val Accuracy: 88.00%
Epoch [17/20], Train Loss: 0.1585, Train Accuracy: 95.86%, Val Loss: 0.3334, Val Accuracy: 88.00%
Epoch [18/20], Train Loss: 0.1542, Train Accuracy: 96.00%, Val Loss: 0.3354, Val Accuracy: 88.00%
Epoch [19/20], Train Loss: 0.1537, Train Accuracy: 96.00%, Val Loss: 0.3356, Val Accuracy: 88.00%
Epoch [20/20], Train Loss: 0.1532, Train Accuracy: 96.00%, Val Loss: 0.3356, Val Accuracy: 88.00%
7/7 ━━━━━━ 1s 49ms/step

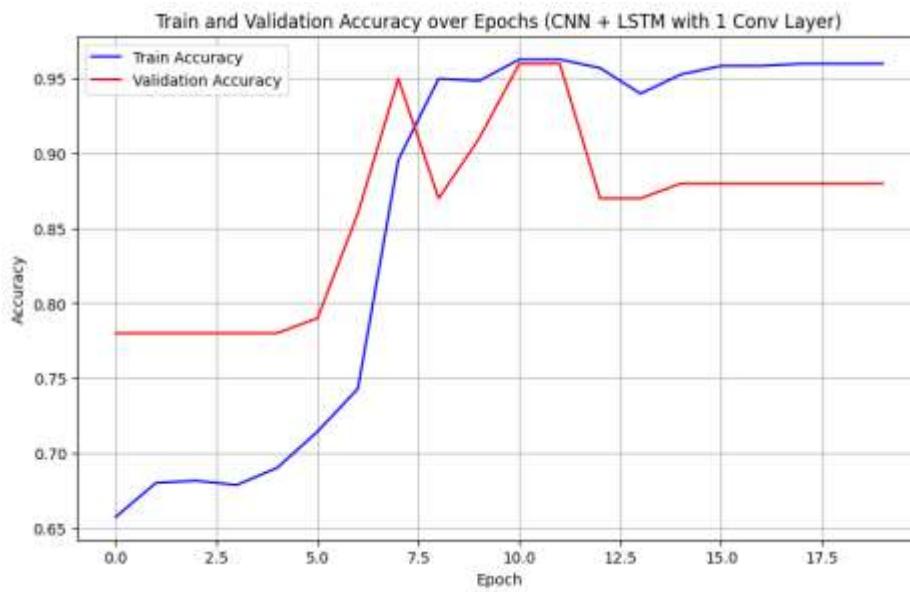
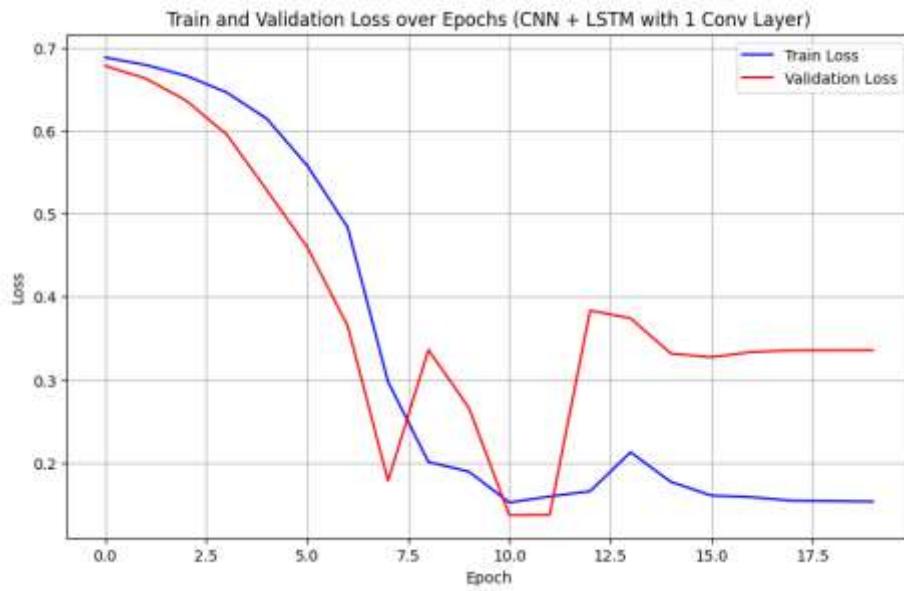
```

Test Metrics for CNN + LSTM (1 Conv layer):
Accuracy: 87.00%
Precision: 100.00%
Recall: 74.00%
F1-Score: 85.06%

مدل ترکیبی CNN + LSTM در طول آموزش روند پیشرفت قابل توجهی را نشان می دهد؛ دقت اعتبارسنجی از ۷۸٪ در دوره های ابتدایی به ۹۶٪ در دوره ۱۱ افزایش می یابد و سپس در بازه ۸۷٪ تا ۸۸٪ تا پایدار می ماند. این پایداری نسبی می تواند ناشی از شروع بیش برازش پس از دوره دوازدهم باشد. عملکرد مدل روی داده های آزمون نیز قابل قبول است: دقت ۸۷٪ همراه با Precision صد درصدی نشان می دهد که مدل در شناسایی نمونه های مثبت بسیار محافظه کار عمل می کند، در حالی که Recall نسبتاً پایین تر (۷۴٪) بیانگر از دست رفتن برخی نمونه های واقعی مثبت است. این نتایج حاکی از تعادل نسبتاً خوب میان اطمینان در پیش بینی و پوشش نمونه های مثبت می باشد.

با رسم نمودارهای خط و دقت برای آموزش و اعتبارسنجی نیز می توانیم تحلیل های بالا را بهتر درک کنیم و می بینیم که خطای آموزش با شبی خوبی در حال کاهش هست نوسانات زیادی ندارد؛ اما برای اعتبارسنجی دچار نوسانات زیادی شده است و حقی بعد از مدتی خطای آن افزایش نیز می یابد. برای دقت هم می بینیم که آموزش مجدد با شبی خوبی افزایش یافته است اما اعتبارسنجی مثل قبل نوسانات زیادی دارد و در نهایت مقدار دقت از آن مقدار ماکسیمم کاهش پیدا ممی کند و پس از مدتی شبی حدودا ثابتی پیدا می کند.

در ادامه نمودارهای خط و دقت را نیز می توانیم برای درک بهتر مشاهده کنیم:



در ادامه از معماری دو لایه پیچشی (Conv1D) به همراه لایه‌های زیرنمونه‌برداری (MaxPooling1D) پیش از LSTM استفاده کردہ‌ایم تا ویژگی‌های محلی متنوعتری از توالی استخراج کنیم. این ترکیب به مدل امکان می‌دهد

الگوهای پیچیده‌تری را پیش از انتقال به حافظه بلندمدت یاد بگیرد و نمای غنی‌تری از داده بسازد. در ادامه، خروجی این لایه‌ها به LSTM و سپس به لایه نهایی متصل شده تا دسته‌بندی دودویی انجام گیرد.

خروجی این مدل برای این 20 اپک و معیارهای محاسبه شده را می‌توانیم در شکل زیر مشاهده کنیم:

▪ 2-layers:

```
Epoch [1/20], Train Loss: 0.6886, Train Accuracy: 58.57%, Val Loss: 0.6774, Val Accuracy: 78.00%
Epoch [2/20], Train Loss: 0.6750, Train Accuracy: 69.43%, Val Loss: 0.6516, Val Accuracy: 78.00%
Epoch [3/20], Train Loss: 0.6479, Train Accuracy: 69.86%, Val Loss: 0.5926, Val Accuracy: 79.00%
Epoch [4/20], Train Loss: 0.5913, Train Accuracy: 70.86%, Val Loss: 0.4955, Val Accuracy: 77.00%
Epoch [5/20], Train Loss: 0.5326, Train Accuracy: 71.57%, Val Loss: 0.4344, Val Accuracy: 82.00%
Epoch [6/20], Train Loss: 0.4122, Train Accuracy: 80.57%, Val Loss: 0.3089, Val Accuracy: 90.00%
Epoch [7/20], Train Loss: 0.2366, Train Accuracy: 93.29%, Val Loss: 0.2273, Val Accuracy: 92.00%
Epoch [8/20], Train Loss: 0.1602, Train Accuracy: 96.14%, Val Loss: 0.1480, Val Accuracy: 96.00%
Epoch [9/20], Train Loss: 0.1544, Train Accuracy: 96.57%, Val Loss: 0.2030, Val Accuracy: 93.00%
Epoch [10/20], Train Loss: 0.1473, Train Accuracy: 96.29%, Val Loss: 0.2390, Val Accuracy: 92.00%
Epoch [11/20], Train Loss: 0.1226, Train Accuracy: 97.14%, Val Loss: 0.1465, Val Accuracy: 96.00%
Epoch [12/20], Train Loss: 0.1104, Train Accuracy: 97.43%, Val Loss: 0.1393, Val Accuracy: 96.00%
Epoch [13/20], Train Loss: 0.1010, Train Accuracy: 97.71%, Val Loss: 0.1129, Val Accuracy: 97.00%
Epoch [14/20], Train Loss: 0.0868, Train Accuracy: 98.14%, Val Loss: 0.2722, Val Accuracy: 91.00%
Epoch [15/20], Train Loss: 0.0876, Train Accuracy: 98.14%, Val Loss: 0.2382, Val Accuracy: 92.00%
Epoch [16/20], Train Loss: 0.0856, Train Accuracy: 98.14%, Val Loss: 0.2245, Val Accuracy: 93.00%
Epoch [17/20], Train Loss: 0.0764, Train Accuracy: 98.43%, Val Loss: 0.2210, Val Accuracy: 93.00%
Epoch [18/20], Train Loss: 0.0755, Train Accuracy: 98.43%, Val Loss: 0.2032, Val Accuracy: 94.00%
Epoch [19/20], Train Loss: 0.0684, Train Accuracy: 98.57%, Val Loss: 0.1192, Val Accuracy: 97.00%
Epoch [20/20], Train Loss: 0.0556, Train Accuracy: 99.00%, Val Loss: 0.1161, Val Accuracy: 97.00%
7/7 ————— 1s 59ms/step

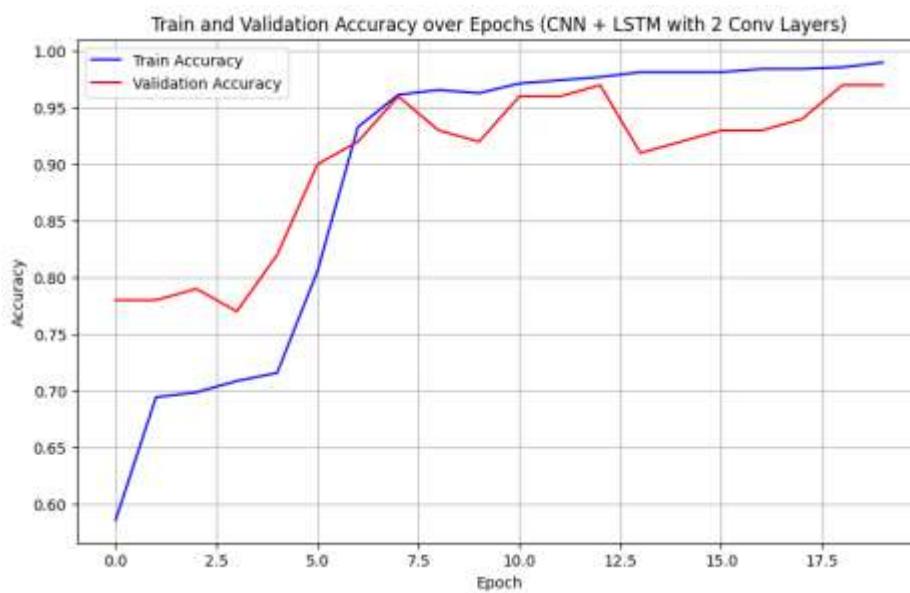
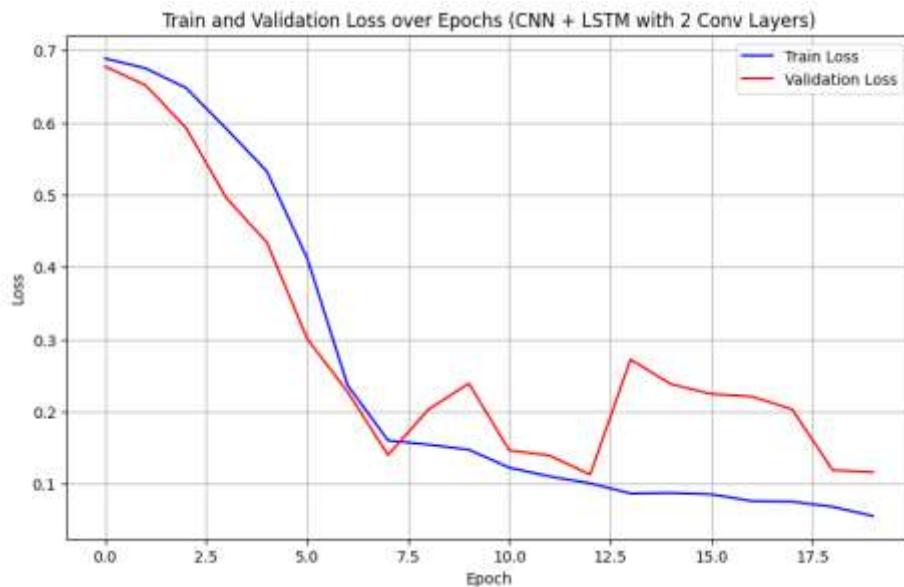
Test Metrics for CNN + LSTM:
Accuracy: 92.50%
Precision: 98.85%
Recall: 98.00%
F1-Score: 91.98%
```

مدل ترکیبی با دو لایه پیچشی عملکرد قابل توجهی از خود نشان داده است. در طول آموزش، شاهد رشد پایدار دقت آموزش از ۵۸.۵۷٪ به ۹۹٪ و کاهش معنادار خطای اعتبارسنجی تا مقدار ۱۱٪ هستیم که نشان دهنده یادگیری مؤثر ویژگی‌ها توسط مدل است. دقت بالای اعتبارسنجی (۹۷٪) و تست (۹۲.۵٪) در کنار مقدار F1 بالا (۹۱.۹۸٪) گواهی بر تعمیم‌پذیری مناسب مدل هستند. همچنین، ترکیب مقدار Precision بسیار بالا (۹۸.۸۵٪) با Recall بالا (۸۶٪) نشان می‌دهد مدل علاوه بر تشخیص دقیق موارد مثبت، توانایی خوبی در شناسایی تعداد زیادی از آن‌ها دارد. این نتایج نشان می‌دهد استفاده از چند لایه پیچشی باعث تقویت استخراج ویژگی و بهبود عملکرد کلی شده است.

با رسم نمودارهای خطای آموزش و اعتبارسنجی نیز می‌توانیم تحلیل‌های بالا را بهتر درک کنیم و می‌بینیم که خطای آموزش با شبی خوبی در حال کاهش هست نوسانات زیادی ندارد؛ اما برای اعتبارسنجی دچار نوسانات زیادی شده است و حقی بعد از مدتی خطای آن افزایش نیز می‌یابد و مجدد مقدارش کم می‌شود اما

نسبت به مدل 1 لایه بهتر عمل کرده است و نوسانات کوتاه تر و کمتری دارد. برای دقیق هم می بینیم که آموزش مجدد با شبیخوبی افزایش یافته است و اعتبارسنجی هم مثل تحلیل قبلی برای 1 لایه است اما اینجا هم شاهد آن هستیم که این نوسانات کمتر و کوتاه تر شده اند.

در ادامه نمودارهای خطای و دقیق را نیز می توانیم برای درک بهتر مشاهده کنیم:



Test Metrics for CNN + LSTM:
Accuracy: 92.50%
Precision: 98.85%
Recall: 86.00%
F1-Score: 91.98%

Test Metrics for CNN + LSTM (1 Conv layer):
Accuracy: 87.00%
Precision: 100.00%
Recall: 74.00%
F1-Score: 85.06%

شکل سمت راست برای 1 لایه و سمت چپ نیز برای 2 لایه می‌باشد. می‌بینیم که دقت در این مدل افزایش قابل توجهی داشته است و نزدیک به 5.5 درصد زیاد شده است؛ همینطور برای سه معیار دیگر هم می‌بینیم که افزایش خیلی خوبی نسبت به مدل سوال 2 داشته است (به جز Precision که حدود 1 درصد مقدارش کمتر شده است). به طور کلی در این مدل توансیتم که معیارهای Accuracy، Recall، Precision و F1-score را بهبود بدهیم و مدل 2 لایه عملکرد خیلی بهتری از خود نشان داد و همینطور اگر به نمودارهای دقت و خطای که برای آموزش و اعتبارسنجی نشان دادیم دقت کنیم، می‌بینیم که در مدل 2 لایه تا حد خوبی از نوسان‌های زیاد نسبت به 1 لایه کاسته شده است و مدل روند نسبتاً خوبی را در پیش بینی ایمیل‌های اسپم و عادی در پیش گرفته است و با دقت کلی 92.5٪ به نتایج بهتری نسبت به مدل سوال یک لایه توanstه است دست یابد.

پس با افزایش تعداد لایه‌ها می‌تواند عملکرد مدل تا حدودی بهتر شود و از بخش‌های قبل دیدیم که مدل LSTM دوطرفه با دقت 95٪ نسبت به باقی مدل‌ها عملکرد بهتر و بالاتری داشت و این مدل 2 لایه در این سوال حدوداً 2.5٪ دقت کمتری نسبت به آن مدل داشته است:

Test Metrics for Bidirectional
Accuracy: 95.50%
Precision: 98.92%
Recall: 92.00%
F1-Score: 95.34%

7

در این بخش با بررسی نمونه‌ای از ایمیل‌های طبقه‌بندی شده توسط مدل، تلاش می‌کنیم موارد موفق و اشتباه را تحلیل کنیم تا علت خطاها مدل را از منظر ویژگی‌های متغیر یا محدودیت‌های ساختاری آن شناسایی نماییم. این تحلیل می‌تواند دید بهتری نسبت به نقاط قوت و ضعیفی مدل ارائه دهد.

در این مرحله با استفاده از مدل دوطرفه (که عملکرد بهتر و بالاتری نسبت به مدل‌های دیگر داشت)، برچسب‌های پیش‌بینی شده برای داده‌های آزمون را استخراج کردیم و سپس نمونه‌هایی از پیش‌بینی‌های درست و نادرست را انتخاب و نمایش دادیم تا بتوانیم رفتار مدل را در مواجهه با متون مختلف تحلیل کنیم.

در شکل زیر به عنوان مثال سه نمونه از پیش‌بینی‌های درست و همچنین سه نمونه از پیش‌بینی‌های نادرست را در خروجی نمایش دادیم و به علت طولانی بودن برخی ایمیل‌های نمایش داده شده، بخشی از متن را در خروجی نمایش دادیم:

- True Predictions:
[Label: 1 - Predicted: 1] موهای ضعیف میباشد عماز ادامه دهد اب ارکش تعاوید محصول شماره مجوز میباشد محصول شماره مجوز میباشد
[Label: 0 - Predicted: 0] گرفتن خوبه هیچیز دیگه باور زمان عقیقی میباشد هرگز علوم تحقیقات تعریفتم لان دیگه نمی رفت نمیشه کاری
[Label: 0 - Predicted: 0] انجعوی دکتری بگیرند دانشکده زبان علوم تحقیقات معوطه کوچکی داره دکتر گلfram میگفتند محیط شنیده دانشگاه
- False Predictions:
[Label: 1 - Predicted: 0] زمزان داروی تازابی باتوان کد کالا قیمت تومان خوب داروی تازابی باتوان طریق فروتول داروی تازابی باتوان
[Label: 1 - Predicted: 0] جالد پیشینی فوتمال جوابت میلیوتی حتما عضو شوید سورت اصولی سایتها فعالیت میتوانید راحتی پولدار شوید
[Label: 0 - Predicted: 1] سازی علیرغم برون سهاره افور عرض عنوان معروف جهت حفظ اعتماد عمومی امولگرایی به ارادتمند علیاصغر لطفی

مدل دوطرفه (LSTM + CNN) با دو لایه کانولوشن، با وجود دستیابی به عملکرد نسبتاً مطلوبی در ارزیابی نهایی همچنان در برخی موارد دچار خطأ شده‌اند. تحلیل دقیق این خطاهای نشان می‌دهد که علل آن را می‌توان در دو دسته کلی قرار داد: ویژگی‌های ذاتی متون و محدودیت‌های ساختاری مدل.

از یک‌سو، بررسی نمونه‌های نادرست طبقه‌بندی شده نشان می‌دهد که برخی از ایمیل‌ها دارای ابهام معنایی بوده یا از واژگان مشترک میان دو کلاس اسپم و عادی استفاده کرده‌اند. برای مثال، ایمیل‌های غیر اسپمی که شامل عباراتی چون «رایگان»، «تحفیف» یا ... هستند، ممکن است به اشتباہ اسپم تشخیص داده شوند. همچنین، کوتاه بودن برخی متون یا ساختار نگارشی غیر متعارف و یا طول خیلی زیاد ایمیل‌ها باعث می‌شود مدل اطلاعات کافی برای تصمیم‌گیری دقیق نداشته باشد. این‌ها نمونه‌هایی از چالش‌های ناشی از پیچیدگی‌های داده هستند که حتی مدل‌های قوی نیز با آن‌ها دست‌وپنجه نرم می‌کنند. از سوی دیگر، با وجود عملکرد بهتر مدل‌های عمیق نسبت به RNN ساده، محدودیت‌هایی نیز در خود معماری مدل‌ها مشاهده می‌شود. مدل‌های LSTM، حتی در حالت دوطرفه، در درک روابط معنایی پیچیده یا مفاهیم ضمنی و چندلایه ناتواناند. همچنین، استفاده از آستانه‌ی پیش‌فرض ۵٪ برای تصمیم‌گیری باعث می‌شود نمونه‌هایی که در مرز بین دو کلاس قرار دارند، دچار برچسب‌گذاری نادرست شوند. علاوه بر این، نسبت به مدل‌های زبانی مدرن نظری BERT که بر پایه ساختارهای Transformer طراحی شده‌اند، این مدل‌ها توانایی کمتری در درک دقیق زبان طبیعی دارند.

در مجموع، می‌توان نتیجه گرفت که خطاهای مدل از ترکیب دو عامل اصلی ناشی می‌شوند: نخست، پیچیدگی و ابهام موجود در برخی متون ایمیل‌ها و شباهت‌های زیاد بین کلاس‌ها؛ دوم، محدودیت‌های معماری و پردازشی مدل‌های مورد استفاده. برای بهبود دقت مدل در مواجهه با این چالش‌ها، می‌توانیم از مدل‌های زبانی پیشرفته‌تر و مواردی که ذکر کردیم استفاده کنیم.