



# مبانی یادگیری ماشین

پاییز ۱۴۰۴

استاد: دکتر حامد ملک

دانشکده مهندسی و علوم کامپیوتر

## پروژه تشخیص احساسات با داده‌های چندوجهی

### هدف پروژه

هدف این پروژه طراحی و پیاده‌سازی یک سیستم تشخیص احساسات است که با بهره‌گیری از یادگیری ماشین و یادگیری عمیق، و با استفاده از دو مدلیتی متن و گفتار، احساس گوینده را پیش‌بینی کند. در این مسئله، هر نمونه داده شامل متن<sup>۱</sup> گفتار<sup>۲</sup> و سیگنال صوتی بوده و مدل باید با ترکیب هر دو، طبقه‌بندی احساسات را انجام دهد. تمرکز پروژه بر ساخت یک خط‌لوله‌ی انتها به انتهای آنها<sup>۳</sup> از پیش‌پردازش و استخراج ویژگی تا آموزش و ارزیابی است، به‌گونه‌ای که مدل نهایی عملکرد قابل قبولی روی داده‌های ارزیابی ارائه دهد.

### جزئیات پروژه

در این پروژه انتظار می‌رود یک مدل تشخیص احساسات بر پایه‌ی داده‌های چندوجهی<sup>۴</sup> ارائه شود. بطوری که برای هر نمونه، متن گفتار و سیگنال صوتی به صورت متناظر در دسترس بوده و خروجی سیستم یک Emotion Label باشد. از نظر پردازش متن، دست شما برای استفاده از مدل‌های از پیش‌آموزش دیده<sup>۵</sup> باز است و می‌توانید بسته به طراحی خود از وزن‌های آماده استفاده کنید، آن‌ها را ثابت<sup>۶</sup> نگه دارید یا با راهبردهای ریزنظمی<sup>۷</sup> برای مسئله‌ی تشخیص احساسات بهینه‌سازی کنید، با این حال لازم است نحوه‌ی استفاده و دلیل انتخاب این راهبردها در گزارش شفاف توضیح داده شود. برای مدلیتی صوت نیز استفاده از انتقال یادگیری<sup>۸</sup> مجاز است، زیرا داده‌های گفتاری معمولاً نسبت به تغییرات گوینده، لهجه، شدت صدا، نویز محیط، کیفیت ضبط و شرایط واقعی حساس‌اند و بهره‌گیری از نمایش‌های آموخته‌شده می‌تواند به تعمیم پذیری<sup>۹</sup> کمک کند. با این وجود، صرفاً تکیه بر یک مدل آماده‌ی «بدون طراحی و بدون آموزش/انطباق» کافی نیست و شما باید معماری کلی و روش‌شناسی پروژه را خودتان طراحی کنید، اجزای لازم برای تطبیق با وظیفه را تعریف کنید و فرآیند آموزش و ارزیابی را به صورت قابل بازتولید<sup>۱۰</sup> ارائه دهید. ارزیابی باید روی تقسیم‌بندی استاندارد داده انجام شود و عملکرد با معیارهای مناسب گزارش و تحلیل گردد، علاوه بر شاخص‌هایی مانند F1 / Accuracy / Precision / Recall ، ارائه‌ی تحلیل‌های تکمیلی مانند Confusion Matrix و تحلیل الگوهای خطای می‌تواند کیفیت گزارش را افزایش دهد.

<sup>1</sup> Speech Transcript

<sup>2</sup> End-to-End Pipeline

<sup>3</sup> Multimodal

<sup>4</sup> Pretrained Language Models

<sup>5</sup> Freezing

<sup>6</sup> Fine-tuning Strategies

<sup>7</sup> Transfer Learning

<sup>8</sup> Robustness / Generalization

<sup>9</sup> Reproducibility

## دادگان

برای این پروژه، دادگان باید به گونه‌ای باشد که برای هر نمونه هر دو مدلیته‌ی متن و صوت در دسترس باشد و علاوه بر آن برچسب احساس نیز ارائه شود؛ به عنوان نمونه، دادگان‌هایی مانند IEMOCAP و MELD این ویژگی را دارند و شامل گفتار به همراه متن متناظر و برچسب‌گذاری احساسی هستند. ارزیابی بر اساس ۷ برچسب احساسی انجام می‌شود، (Anger, Disgust, Fear, Joy, Neutral, Sadness, Surprise) و بنابراین دادگان انتخابی شما نیز باید همین برچسب‌ها را داشته باشد یا امکان نگاشت<sup>۱۰</sup> دقیق آن‌ها به این مجموعه فراهم باشد. همچنین باید توجه داشت که در مرحله‌ی تحويل، مدل روی نمونه‌های آزمون تصادفی بررسی می‌شود. بنابراین انتخاب دادگان و آماده‌سازی آن باید به گونه‌ای انجام شود که مدل نهایی تعیین‌پذیری و پایداری مناسبی در برابر تغییرات واقعی مانند تفاوت گوینده و شرایط ضبط داشته باشد، و در گزارش نیز اختلاف تعریف احساسات و عدم توازن کلاس‌ها<sup>۱۱</sup> بین دادگان‌ها شفاف توضیح داده شود.

## ملاحظات اجرایی و تحويل پروژه

در تحويل نهایی، انتظار می‌رود تمامی اجزای پروژه به صورت کامل و قابل اجرا ارائه شوند، شامل کدها، فایل‌های تنظیمات و مستندات لازم برای بازتولید نتایج. گزارش پروژه باید تصمیمات طراحی، روش‌شناسی، معماری مدل، فرایند آموزش و تنظیم ابرپارامترها<sup>۱۲</sup> و نتایج ارزیابی را به صورت شفاف پوشش دهد و علاوه بر گزارش معیارهای اصلی، تحلیل خطا و تفسیر نتایج را نیز ارائه کند.

## موارد امتیازی

۱. ارائه‌ی یک واسط سبک برای تست مدل مانند UI ساده یا API جهت بارگذاری ورودی و مشاهده خروجی
۲. اجرای اعتبارسنجی متقابل<sup>۱۳</sup> برای گزارش پایدارتر نتایج.
۳. انجام Ablation Study کامل؛ مقایسه‌ی فقط متن، فقط صوت، و ترکیبی و تحلیل اثر هر کدام
۴. دستیابی به عملکرد بالاتر از یک آستانه‌ی مشخص و دریافت امتیاز وزن‌دار مبتنی بر رتبه‌بندی.

سالم و موفق باشید.

<sup>10</sup> Label Mapping

<sup>11</sup> Class Imbalance

<sup>12</sup> Hyperparameters

<sup>13</sup> Cross-Validation