Model Index	Model Ref	Model Name	Model Name	Version	Context Length	Min	Sec	Latency (s / record) [1]	Input Token Price ($ / 1K tokens)	Input Tokens	Output Token Price ($ / 1K tokens)	Output Tokens	Total Price ($ = Input + Output)	Avg Price ($ / record)	Num Records	TP	TN	FP	FN	AP	AN	Precision	Recall	F1	Num Manual Clean	Notes
00			Gemini 1.5 Flash	Original				--														0.3555	0.6646	0.4632		The first baseline produced on 12/13/2024 with the dataset.
01			Gemini 1.5 Flash	Baseline				--														0.6369	0.6646	0.6505		Improved from the 00 version by extracting better excerpts & revised prompt.
0	0.haiku3.5_baseline_eval.csv	anthropic.claude-3-5-haiku-20241022-v1:0	Claude 3.5 Haiku	Baseline	200k	24	3	0.4234	$0.0008	2,268,354	$0.0040	249,544	$2.81	$0.00	1084	114	784	151	35	149	935	0.4302	0.7651	0.5507	20	
1	1.haiku3.5_v213_prompt_eval.csv	anthropic.claude-3-5-haiku-20241022-v1:0	Claude 3.5 Haiku	v213 prompt	200k	24	38	0.4557	$0.0008	4,094,894	$0.0040	252,836	$4.29	$0.00	1084	112	868	67	37	149	935	0.6257	0.7517	0.6829	1	
2	2.sonnet3.5v2_v213_prompt_eval.csv	anthropic.claude-3-5-sonnet-20241022-v2:0	Claude 3.5 Sonnet	v213 prompt	200k	15	26	0.2869	$0.0030	4,094,894	$0.0150	126,942	$14.19	$0.01	1084	117	891	44	32	149	935	0.7267	0.7852	0.7548	0	
3	3.sonnet3.5_v214_prompt_eval.csv	anthropic.claude-3-5-sonnet-20241022-v2:0	Claude 3.5 Sonnet	v214 prompt	200k	31	53	0.5923	$0.0030	4,204,378	$0.0150	347,006	$17.82	$0.02	1084	112	890	45	37	149	935	0.7134	0.7517	0.7320	0	
4	4.sonnet3.5_v214_2_prompt_eval.csv	anthropic.claude-3-5-sonnet-20241022-v2:0	Claude 3.5 Sonnet	v214_v2 prompt	200k	15	16	0.2777	$0.0030	4,051,534	$0.0150	132,559	$14.14	$0.01	1084	118	889	46	31	149	935	0.7195	0.7919	0.7540	0	
5	5.sonnet3.5v2_v217_prompt_eval.csv	anthropic.claude-3-5-sonnet-20241022-v2:0	Claude 3.5 Sonnet	v217 prompt	200k	14	42	0.2841	$0.0030	3,970,234	$0.0150	129,905	$13.86	$0.01	1084	113	900	35	36	149	935	0.7635	0.7584	0.7609	1	
6	6.sonnet3.7_v213_prompt_eval.csv	us.anthropic.claude-3-7-sonnet-20250219-v1:0	Claude 3.7 Sonnet	v213 prompt	200k	23	2	0.4050	$0.0030	4,097,062	$0.0150	158,999	$14.68	$0.01	1084	123	875	60	26	149	935	0.6721	0.8255	0.7410	7	
7	7.sonnet3.7_v217_prompt_eval.csv	us.anthropic.claude-3-7-sonnet-20250219-v1:0	Claude 3.7 Sonnet	v217 prompt	200k	21	37	0.4022	$0.0030	3,972,402	$0.0150	146,084	$14.11	$0.01	1084	125	850	85	24	149	935	0.5952	0.8389	0.6964	2	
8	0.commandR_v213_prompt_eval.csv	cohere.command-r-v1:0	Cohere Command R	v213 prompt	128k	11	72	0.2592	$0.0005	3,871,119	$0.0015	93,241	$2.08	$0.00	1084	112	650	285	37	149	935	0.2821	0.7517	0.4103	52	
9	1.commandR+_v213_prompt_eval.csv	cohere.command-r+	Cohere Command R+	v213 prompt	128k	12	36	0.2435	$0.0030	3,871,119	$0.0150	79,691	$12.81	$0.01	1084	84	883	52	65	149	935	0.6176	0.5638	0.5895	5	
10	2.commandR+_baseline_eval.csv	cohere command R+	Cohere Command R+	Baseline	128k	13	5	0.2325	$0.0030	2,211,515	$0.0150	110,233	$8.29	$0.01	1084	100	808	127	49	149	935	0.4405	0.6711	0.5319	1	
11	0.llama31_405B_v213_prompt_eval.csv	meta.llama3-1-405b-instruct-v1:0	Llama 3.1 405B Instruct	v213 prompt	128k	29	54	0.5581	$0.0024	3,451,260	$0.0024	111,363	$8.55	$0.01	1084	127	856	79	22	149	935	0.6165	0.8523	0.7155	0	
12	1.llama31_405B_v217_prompt_eval.csv	meta.llama3-1-405b-instruct-v1:0	Llama 3.1 405B Instruct	v217 prompt	128k	30	57	0.5784	$0.0024	3,582,424	$0.0024	115,951	$8.88	$0.01	1084	128	851	84	21	149	935	0.6038	0.8591	0.7091	6	
13	2.llama31_405B_baseline_eval.csv	meta.llama3-1-405b-instruct-v1:0	Llama 3.1 405B Instruct	Baseline	128k	35	42	0.6522	$0.0024	2,073,496	$0.0024	207,453	$5.47	$0.01	1084	123	858	77	26	149	935	0.6150	0.8255	0.7049	20	
14	3.llama32_3B_v213_prompt_eval.csv	meta.llama3-2-3b-instruct-v1:0	Llama 3.2 3B Instruct	v213 prompt	128k	7	21	0.1421	$0.0002	3,472,940	$0.0002	110,578	$0.54	$0.00	1084	60	799	136	89	149	935	0.3061	0.4027	0.3478	6	
15	4.llama33_70B_v213_prompt_eval.csv	meta.llama3-3-70b-instruct-v1:0	Llama 3.3 70B Instruct	v213 prompt	128k	7	16	0.1375	$0.0007	3,474,024	$0.0007	113,929	$2.58	$0.00	1084	130	841	94	19	149	935	0.5804	0.8725	0.6971	1	
16	5.llama33_70B_v217_prompt_eval.csv	meta.llama3-3-70b-instruct-v1:0	Llama 3.3 70B Instruct	v217 prompt	128k	7	36	0.1559	$0.0007	3,606,272	$0.0007	115,983	$2.68	$0.00	1084	127	845	90	22	149	935	0.5853	0.8523	0.6940	2	
17	0.mistral_7B_v213_prompt_eval.csv	mistral.mistral-7b-instruct-v0:2	Mistral 7B Instruct	v213 prompt	32k	12	38	0.2454	$0.0002	4,057,785	$0.0002	115,048	$0.63	$0.00	1084	54	829	106	95	149	935	0.3375	0.3624	0.3495	161	
18	1.mistral_8x7B_v213_prompt_eval.csv	mistral.mistral-8x7b-instruct-v0:1	Mistral 8x7B Instruct	v213 prompt	32k	12	67	0.2721	$0.0005	4,047,350	$0.0007	101,617	$1.89	$0.00	1084	71	807	128	78	149	935	0.3568	0.4765	0.4080	34	
19	2.mistral_large_v213_prompt_eval.csv	mistral.mistral-large-2407-v1:0	Mistral Large 2 (24.07)	v213 prompt	131k	14	55	0.2961	$0.0020	4,052,466	$0.0060	123,357	$8.85	$0.01	1084	90	905	30	59	149	935	0.7500	0.6040	0.6691	1	
20	0.nova_micro_v213_prompt_eval.csv	amazon.nova-micro-v1:0	Amazon Nova Micro	v213 prompt	128k	4	50	0.1162	$0.0000	3,576,150	$0.0001	101,895	$0.14	$0.00	1084	102	699	236	47	149	935	0.3018	0.6846	0.4189	0	
21	1.nova_lite_v213_prompt_eval.csv	amazon.nova-lite-v1:0	Amazon Nova Lite	v213 prompt	300k	5	28	0.1135	$0.0001	3,576,150	$0.0002	98,374	$0.24	$0.00	1084	88	822	113	61	149	935	0.4378	0.5906	0.5029	0	
22	2.nova_pro_v213_prompt_eval.csv	amazon.nova-pro-v1:0	Amazon Nova Pro	v213 prompt	300k	7	53	0.1716	$0.0008	3,576,150	$0.0032	104,504	$3.20	$0.00	1084	104	818	117	45	149	935	0.4706	0.6980	0.5622	3	

Note:
1. all models use the same dataset. Other than the Original version, all models use the same excerpts as the context for the classification.
2. parameters for experiments: model provider/version & prompt.

[1] ThreadPool on 8 Core Apple M2 with 5 max workers