

Table 1: Comparison of Performance on Additional Metrics with FeatLLM.

dataset	shot	acc		precision		recall		f1_score	
		FeatLLM	ProtoLLM	FeatLLM	ProtoLLM	FeatLLM	ProtoLLM	FeatLLM	ProtoLLM
Adult	4	44.82±24.99	<b>56.24±5.04</b>	63.36±16.94	<b>66.84±1.09</b>	57.89±9.49	<b>70.35±2.85</b>	38.77±21.95	<b>55.77±4.65</b>
	8	48.24±20.09	<b>58.75±4.84</b>	53.38±24.08	<b>67.33±1.08</b>	64.52±11.75	<b>71.75±2.66</b>	45.69±21.04	<b>58.05±4.35</b>
	16	<b>64.95±18.48</b>	62.86±4.58	<b>71.42±6.46</b>	68.25±1.13	68.34±9.21	<b>73.97±2.42</b>	58.84±15.99	<b>61.70±4.02</b>
Bank	4	32.42±13.68	<b>73.21±6.48</b>	55.40±1.12	<b>61.31±1.86</b>	58.32±4.73	<b>72.55±3.60</b>	30.66±10.73	<b>60.72±4.08</b>
	8	38.50±12.31	<b>75.26±3.93</b>	55.72±0.76	<b>61.72±1.69</b>	60.57±4.47	<b>73.04±3.09</b>	35.69±10.40	<b>62.17±2.82</b>
	16	45.21±10.61	<b>76.67±3.83</b>	56.35±0.96	<b>62.92±1.38</b>	63.57±3.65	<b>75.07±1.75</b>	41.25±7.75	<b>63.92±2.65</b>
Blood	4	62.94±7.12	<b>63.11±9.33</b>	60.73±3.94	<b>62.86±5.37</b>	63.31±4.29	<b>65.55±6.81</b>	58.55±5.35	<b>58.74±7.57</b>
	8	66.36±4.89	<b>64.98±6.98</b>	60.85±3.39	<b>63.60±5.10</b>	62.96±5.46	<b>67.67±6.42</b>	59.31±3.92	<b>61.25±6.58</b>
	16	64.06±8.43	<b>65.29±4.88</b>	60.60±3.78	<b>63.20±3.52</b>	62.21±4.73	<b>67.43±4.43</b>	57.85±6.10	<b>61.65±4.38</b>
Credit-g	4	64.97±4.43	<b>69.97±2.97</b>	51.13±7.60	<b>64.35±10.74</b>	51.01±2.55	<b>51.77±1.45</b>	47.33±4.63	46.23±2.84
	8	64.13±4.97	<b>69.90±3.32</b>	52.76±5.10	<b>63.59±9.91</b>	51.59±3.15	<b>52.72±1.93</b>	50.39±3.59	48.56±3.79
	16	63.43±4.68	<b>69.67±3.04</b>	55.71±6.59	<b>61.11±3.87</b>	54.39±5.12	<b>55.48±2.32</b>	53.03±5.84	<b>54.23±3.45</b>

Table 2: Performance of Weighted ProtoLLM and Weighted ProtoLLM with Half of the Features Dropped.

dataset	shot	Weighted-protoLLM	Weighted-protoLLM(del features)
adult	4	85.34	85.32
	8	85.5	85.45
	16	85.72	85.63
bank	4	81.01	80.4
	8	81.01	80.35
	16	81.29	80.63
blood	4	71.34	69.59
	8	71.46	69.19
	16	72.85	70.26
diabetes	4	80.73	80.42
	8	80.79	80.47
	16	80.72	80.29
NHANES	4	95.95	96.43
	8	96.57	96.92
	16	98.37	98.47

Table 3: Performance of ProtoLLM with Poor or Missing Feature Descriptions.

dataset	shot	full-desc	poor-desc	no-desc
Adult	4	86.01±0.78	83.36±0.67	83.63±0.55
	8	86.12±0.92	83.68±0.67	83.90±0.59
	16	86.28±0.77	83.91±0.82	84.08±0.81
Bank	4	80.85±2.58	71.99±1.93	61.15±4.91
	8	81.41±2.58	72.98±1.90	62.86±4.98
	16	83.26±1.40	75.39±1.65	65.53±4.67
Blood	4	75.98±4.99	72.78±7.12	62.44±9.80
	8	76.35±4.61	74.69±6.58	64.18±8.88
	16	75.46±4.12	74.88±4.75	66.54±5.84
Credit-g	4	62.25±2.86	51.32±4.75	51.39±4.85
	8	63.26±2.87	52.34±4.54	52.74±4.27
	16	64.52±3.28	54.59±4.66	55.48±4.73