RFormula

- Linguagem R permite definir modelo através de fórmula
- [variável dependente] ~ [variável independentes]
- Variáveis Independentes podem ser definidas através de +
- Ponto define todas os atributos – variável dependente
- Spark implemente Rformula
  - Combina variáveis independentes em uma única coluna

Exemplo

```
HP ~ Consumo + Cilindros+ Cilindradas

HP ~ .
```

RFormula

- Spark implementa Rformula
  - Colunas numéricas serão transformadas em double
  - Strings serão transformadas com StringIndexer, e a última categoria é excluída e então aplica One HotEncoding

Consumo	Cilindros	Cilindradas	RelEixoTraseiro	Peso	Tempo	TipoMotor	Transmissao	Marchas	Carburadors	HP
21	6	160	39	262	1646	0	1	4	4	110
21	6	160	39	2875	1702	0	1	4	4	110
228	4	108	385	232	1861	1	1	4	1	93
214	6	258	308	3215	1944	1	0	3	1	110
187	8	360	315	344	1702	0	0	3	2	175
181	6	225	276	346	2022	1	0	3	1	105
143	8	360	321	357	1584	0	0	3	4	245
244	4	1467	369	319	20	1	0	4	2	62
228	4	1408	392	315	229	1	0	4	2	95
192	6	1676	392	344	183	1	0	4	4	123
178	6	1676	392	344	189	1	0	4	4	123
164	8	2758	307	407	174	0	0	3	3	180
173	8	2758	307	373	176	0	0	3	3	180
152	8	2758	307	378	18	0	0	3	3	180
104	8	472	293	525	1798	0	0	3	4	205
104	8	460	3	5424	1782	0	0	3	4	215
147	8	440	323	5345	1742	0	0	3	4	230
324	4	787	408	22	1947	1	1	4	1	66
304	4	757	493	1615	1852	1	1	4	2	52
339	4	711	422	1835	199	1	1	4	1	65

- HP ~ Consumo + Cilindros + Cilindradas"

independente	dependente
[21.0,6.0,160.0]	110.0
[21.0,6.0,160.0]	110.0
[228.0,4.0,108.0]	93.0
[214.0,6.0,258.0]	110.0
[187.0,8.0,360.0]	175.0
[181.0,6.0,225.0]	105.0
[143.0,8.0,360.0]	245.0
[244.0,4.0,1467.0]	62.0
[228.0,4.0,1408.0]	95.0
[192.0,6.0,1676.0]	123.0
[178.0,6.0,1676.0]	123.0
[164.0,8.0,2758.0]	180.0
[173.0,8.0,2758.0]	180.0
[152.0,8.0,2758.0]	180.0
[104.0,8.0,472.0]	205.0
[104.0,8.0,460.0]	215.0
[147.0,8.0,440.0]	230.0
[324.0,4.0,787.0]	66.0
[304.0,4.0,757.0]	52.0
[339.0,4.0,711.0]	65.0