We develop a method to train a document embedding model with an unlabeled dataset and low computational resources. Using teacher-student training, we distill SBERT's capacity to capture text structure and Paragraph Vector's ability to encode extended context into the resulting embedding model. We test our method on Longformer, a Transformer model with sparse attention that can process up to 4096 tokens. We explore several loss functions for the distillation of knowledge from the two teachers (SBERT and Paragraph Vector) to our student model (Longformer). Throughout experimentation, we show that despite SBERT's short maximum context, its distillation is more critical to the student's performance. However, the student model can benefit from both teachers. Our method improves Longformer's performance on eight downstream tasks, including citation prediction, plagiarism detection, and similarity search. Our method shows exceptional performance with few finetuning data available, where the trained student model outperforms both teacher models. By showing consistent performance of differently configured student models, we demonstrate our method's robustness to various changes and suggest areas for future work.